TECHNICAL BASIS FOR UTAH'S NUTRIENT STRATEGY

10/10/2014

Development of Stressor-Response Models for Utah Streams

This report documents the results of a stressor-response analysis that Utah's Division of Water Quality conducted to evaluate relationships among functional and ecological responses and nutrient gradients across Utah streams. The indicators developed through these investigations will support the development of numeric nutrient criteria development and nutrient-specific assessment methods. Together these tools will help identify and address cultural eutrophication problems in Utah.

FORWARD: PLANNED EDITS

As we transitioned from this original document toward a formal proposal for headwater criteria, there was additional discussion among the scientists on the Technical Review Team and UDWQ staff. One consequence of these discussions was an agreement to the following edits that are currently in progress:

- Reword Section 2, Chapter 6 (*Structural Indicators: Relationships between Nutrients and Stream Biota*) to remove the TITAN analysis that was conducted on the macroinvertebrate assemblage.
- Rework Chapter 8 (*Summary of Stressor-Response Indicators*) so that it can stand alone as a summary of the technical underpinnings of the analyses that are most directly linked to the headwater criteria proposal.
- Rework Section 3, especially Chapter 10 (*Ambient Nitrogen and Phosphorus in Headwater Streams*) to incorporate the new analyses that limited the headwater distribution data to summertime average calculations.
- Incorporate the Appendix for Chapter 11 (*Accounting for Complexity, Uncertainty, Variability and Covariables in Site-specific Analyses: the path Forward*).
- Add new DO analysis to Chapter 3 (*Stream Metabolism*), but make it clear that the 30-day average DO graph is not based on 30-days of data, so it is a conservative method.
- Look at data to determine if chl-a/AFDM relate to GPP or ER.
    - This was done in the organic matter chapter to some degree, but can we do this more directly?
    - Consider including this in the rewrite of Chapter 8 (see above).
- Add a literature review to that thoroughly explores that various scientists have tried to quantify what constitutes "nuisance" benthic algae concentrations.
- Add an SOP for quantitative visual assessments of filamentous algae to the report *Appendix*.

Technical Basis for Utah's Nutrient Strategy

DEVELOPMENT OF STRESSOR-RESPONSE MODELS FOR UTAH STREAMS

2014

Authors

J. D. Ostermiller
M. Shupryt
M. A. Baker
B. Neilson[2]
E. B. Gaddis
A J. Hobson[2]
B. Marshall[1]
T. Miller[1]
D. Richards[1]
N. von Stackelberg

1. Primary authors for Chapter 10
2. Primary authors for Chapter 11

Technical Review

T. Bosteels
E. Flemer
T. Laidlaw
D. Sorenson
C. Walker

.

Acknowledgements
The authors wish to thank the many individuals who assisted with this research. Particularly Utah's Water Quality Board who provided project funding. The Division of Water Quality's Monitoring Section did a miraculous job collecting the information this report; their many long hours and devotion to the project are greatly appreciated. The Nutrients Workgroup within the Division of Water Quality and the Nutrient Core Team provided valuable input into preliminary results. Walt Baker and Leah Ann Lamb saw the vision and importance of this project and were supportive throughout.

EXECUTIVE SUMMARY

Introduction

Worldwide, humans continue to add excessive levels of nutrients, particularly nitrogen (N) and phosphorus (P), to waterbodies. These inputs have created what many, including Utah's Division of Water Quality (UDWQ) and the United States Environmental Protection Agency (USEPA), consider to be among the most significant threats to water quality. The response to these concerns has been a nationwide effort to reduce inputs of human-caused nutrients to waterbodies. One important component to these efforts is the development of Numeric Nutrient Criteria (NNC), which define N and P concentrations that cannot be exceeded to maintain the health of streams and lakes.

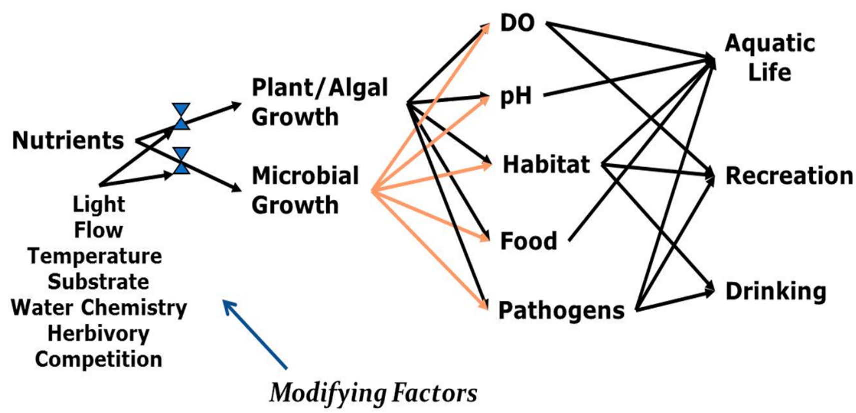


Figure ES1. A simplified conceptual model linking nutrients to aquatic life uses. Nutrients, modified by physical, chemical and ecological factors, influence the composition and abundance of plants and algae (autotrophs) and microbes and fungus (heterotrophs). These assemblages, in turn, alter several characteristics that each have the potential to degrade several aquatic life uses. Not depicted are specific interim processes and important interactions among these responses.

Nutrients act through many interrelated paths that can lead to the degradation of Aquatic Life, Drinking Water, or Recreation Uses (Figure E1). These paths between nutrients and uses are not independent. Moreover, they are moderated by locally divergent physical and chemical processes. From a technical basis, the complexity of these relationships remains a central challenge in address nutrient-related water quality problems. Other challenges are socioeconomic with two primary sources. First, there are many different sources of nutrients and therefore a broad diversity of stakeholders who often have conflicting interests. Second, reducing nutrient inputs to streams and lakes can be costly.

The primary objective of this report is to provide the technical basis for the development of numeric nutrient criteria for Utah's rivers and streams.

To address the complexity of cultural eutrophication problems UDWQ has adapted an Adaptive Management framework. This approach encourages iterative solutions that can be applied while uncertainties that result from the complexity of the problems are resolved. In Utah, numerous streams and lakes have been listed as impaired due to violations of water quality parameters associated with excess nutrients (e.g., Dissolved Oxygen, pH, TSIs). In response to these impairments, several TMDLs have been developed resulting in nutrient management plans for watersheds throughout the state. However, these reactive approaches are insufficient because they too frequently rely on impairment to occur before action is taken. As a result UDWQ, in collaboration with key stakeholders (Nutrient Core Team), has proposed initial solutions that seek affordable reductions from both point- and nonpoint sources. In addition, the group proposes an iterative development of regulations with the promulgation of numeric nutrient criteria (NNC) in headwater streams, to be followed by site-specific criteria development for streams elsewhere. Chapter 1 provides a more thorough background into nutrient-related problems and Utah's proposed solutions.

The primary objective of the research described in this report is to establish the technical basis for the development of NNC to protect aquatic life and recreation uses. To protect aquatic life uses several ecological responses known to be sensitive to nutrient enrichment were identified from the scientific literature. These responses fall within two broad classes: 1) functional responses that can be used quantify changes in important processes or states (Section 2) and 2) structural responses that quantify changes in the composition and abundance of stream biota (Section 2). In both cases, Utah specific thresholds were empirically derived to determine concentrations of N or P that are broadly associated with stream condition—as measured by each response. Generally, two thresholds were derived for each indicator: one to distinguish between streams in good vs. fair condition and one to distinguish between streams in fair vs. poor condition. To protect recreation uses, a survey was conducted that examined the influence of excess algae growth on stream recreation activities (Chapter 7). Once established, the ecological relevance of these statistical thresholds was evaluated with independent measures of stream condition and a review of primary scientific literature. Figure E2 summarizes all aquatic life thresholds obtained from these approaches.

The intended application of the ecological response thresholds is place dependent. For headwaters, the empirical thresholds described in the first two sections of the report will provide multiple lines of evidence that can be used by UDWQ to derive NNC for N and P that are appropriately protective of aquatic life uses. For streams outside of headwaters the thresholds will be used to more accurately identify streams with nutrient-related problems. Moreover, as streams with nutrient related problems are identified, the responses described in this report will provide useful diagnostic information because each indicator describes a different response to excess nutrients. These diagnostic data will help inform study designs for the derivation of site-specific standards and specific remediation actions that are most likely to restore stream conditions.

Ecological Responses

Different approaches were required for the derivation of functional and structural response thresholds. In the case of structural responses, UDWQ has an established biological assessment program, so existing sources of data could be used to determine the concentrations of N or P that were associated with the greatest changes in the composition of macroinvertebrates and diatoms (Chapter 6). However, existing data were not available for functional indicators, so UDWQ established a pilot study to generate measures of these responses from 15 reference sites and then 20 sites that varied in extent of nutrient enrichment (Chapter 2 provides study design details). While these indicators do not capture all possible responses, collectively they provide quantitative measures of all important causal paths between nutrients and responses. Also, the selected indicators are pragmatic because they either capitalize on responses derived from routine and ongoing monitoring, or are dependent on data that could be incorporated into ongoing monitoring efforts with minimal additional resources.

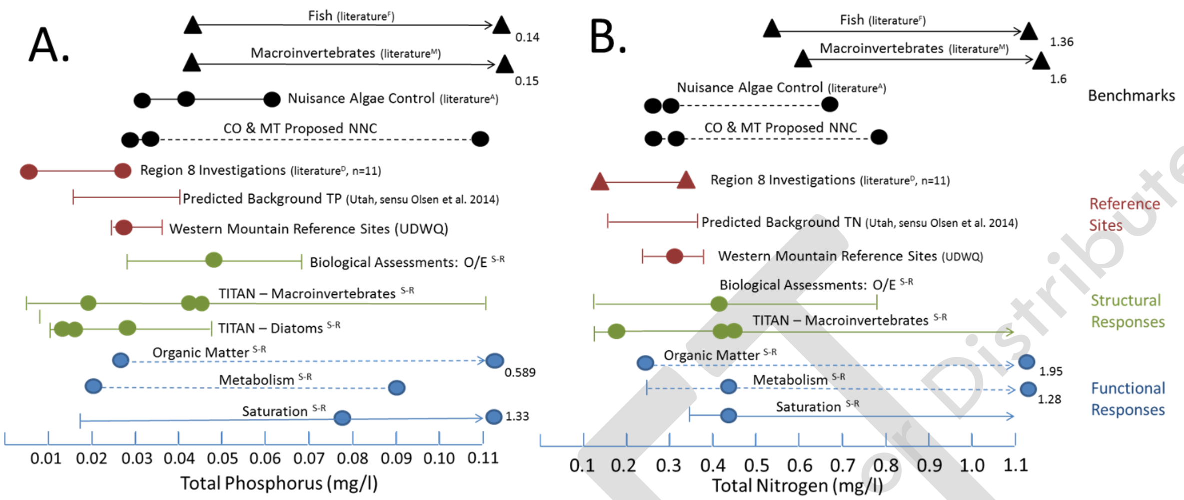


Figure ES2. Threshold established from total phosphorus (panel A.) and total nitrogen (panel B). Dots indicate specific thresholds that we calculated or derived from literature. Solid lines indicates known error (confidence intervals) surrounding thresholds estimates. Lines bracketed by triangles indicates the presence of numerous intermediate thresholds (dots). Stating from bottom, blue text and graphic depict thresholds derived from functional response (Section 1 of this report), whereas green text and graphics depict thresholds for structural responses (Section 2). Red text and graphics indicate thresholds obtained using distribution approaches starting with those obtained from the distribution of headwater reference sites (bottom), then those derived from models and finally those from other published investigations (top). Black text and graphics depict thresholds from other investigations staring with proposed NNC within USEAPA Region 8 (bottom), followed by thresholds obtained from the primary literature.

Functional Responses

Nutrient diffusing substrates (NDS, Chapter 3) were used to provide quantitative estimates of two important states: the nutrient that primarily limits algae production and the concentration of N or P that is associated with saturation—the point where additional nutrient increases do not result in increases in primary production. NDSs are bioassays where growth media are augmented with N, P, or N and P. These experiments are then deployed in streams and the accrual of algae on treatments is compared with controls. Overall, these investigations highlight the importance of considering both N and P in Utah's Nutrient Reduction Strategy. Generally, colimitation of both N and P was the most common condition among study streams, a finding that is consistent with recent literature. Among streams that were not saturated with nutrients, the addition of both N and P in these bioassays resulted in a greater response that the addition of either nutrient alone, which suggests that in streams with excess algae growth the reduction of both N and P is more likely to improve conditions than reductions of either nutrient alone. Thresholds calculated with these data revealed a TP saturation threshold of 0.078 mg/L (±0.017-1.33), and a threshold of 0.42 mg/L (±0.33-1.4) for TN (Figure ES2). In the context of NNC development, this response is likely too extreme to be considered protective because other important ecological responses general occur well before saturation levels. Nevertheless, these thresholds provide important context and have implications with respect to expectations for future nutrient reduction efforts.

Whole stream metabolism techniques were used to obtain measures of two fundamental ecological functions: Gross Primary Production (GPP) which measures rates of primary production via concurrent oxygen production, and Ecosystem Respiration (ER) which measures the growth of animals as utilization of carbon, and concurrent consumption of oxygen, by animal and microbes (Chapter 4). Thresholds were determined to determine TN or TP concentrations that were associated with streams classified as good, fair or poor condition based on measured GPP and ER rates. These calculations suggest that, on average, streams move from good to fair condition at a TP above 0.02 mg/L or a TN of 0.09 mg/L. Similarly, streams generally move from fair to poor condition once TP exceeds 0.09 mg/L or TN exceeds 1.28 mg/L (Figure ES2). Thresholds were also calculated for GPP and ER so that these metrics could be used to assess stream conditions. In general, streams are in poor condition once GPP exceeds 10 or ER exceeds -9 (g O^2/m^2/day), however the confidence is this assertion is lower for streams with low slopes or high canopy cover. Metabolism response thresholds were then compared against independently derived numeric criteria for Dissolved Oxygen (DO). Streams with DO observations below numeric criteria occurred at sites with high GPP and ER. The empirically derived thresholds for GPP and ER from Utah streams were consistent with deleterious responses reported in the primary literature.

The amount of storage and processing of carbon is another important ecosystem function of direct relevance to cultural eutrophication. On one hand, excessive GPP can lead to excessive algae growth that degrades aquatic life uses via diminished quality of habitat or food, or recreation uses via diminished aesthetics. On the other hand, the consumption (i.e., ER) of excessive carbon within a system—both from plants/algae growth and outside sources—can result in low DO, particularly when ambient N or P is high. This study captured the importance of carbon with reach-scale measures of organic matter standing stocks in the stream (Chapter 5). Specifically, thresholds for TN and TP were derived for sources of carbon within the stream that are most strongly related to primary production or most readily available to stream fungus and microbes (e.g., Fine Particulate Organic Matter (FPOM)). Streams with high levels of these sources of organic matter also had higher concentrations of TN and TP. Statistical relationships further suggest that when stream TP exceeds 0.026 mg/L or TN exceeds 0.238 mg/L sites move from good to fair condition, whereas concentration of 0.589 and 1.95 mg/L (TP and TN respectively) distinguish between streams in fair vs. poor condition (Figure ES2). Organic matters standing stock thresholds were also calculated which suggested that streams with stores >48 g AFDM/m^2 are at increased risk of degrading uses. These thresholds were confirmed against independent DO criteria and sites where standing stocks above these thresholds were much more likely to have DO observations below criteria.

Structural Responses

Excess nutrients alter ecological functions, but ultimately it is important to know the extent to which these changes alter stream biota. The functional indicator pilot data were augmented with existing biological assessment data to evaluate relationships between increasing stream nutrients and alterations to the presence and relative abundance macroinvertebrate and diatom assemblages with a statistical technique called TITAN (Chapter 6). Historic laboratory procedures limited the analysis of <u>both</u> assemblages to TP because TN was available at an insufficient number of sites with diatom taxa to conduct the analyses. TITAN uses changes in the presence/absence and relative abundance of taxa within each assemblage to derive three thresholds, one that identifies the TN or TP concentration that best identifies losses of sensitive taxa, one that identifies concentrations associated with increases in tolerant taxa, and an overall threshold that identified the concentration where both groups exhibit the strongest changes. All three thresholds were calculated and report, but this summary is limited to the latter (average) thresholds. For TP, a threshold of ~0.02 mg/L was established for both diatoms (0.022 mg/L ± 0.010-0.047) and macroinvertebrates (0.015 mg/L±0.004-0.113 5th/95th Confidence Intervals) assemblages. A TN threshold of 0.41 mg/L (±0.40-1.1) identified the concentration that, on average, was associated with greatest changes in the composition of macroinvertebrates. For macroinvertebrates thresholds for TN and TP were compared against independently derived biological impairment thresholds (O/E >0.78 or 0.83). On average, streams that were predetermined to be impaired from biological assessments had TN or TP concentrations above the thresholds established for TITAN. The thresholds for both diatoms and macroinvertebrates were also consistent, albeit somewhat lower, than those obtained from similar evaluations elsewhere (Figure ES2).

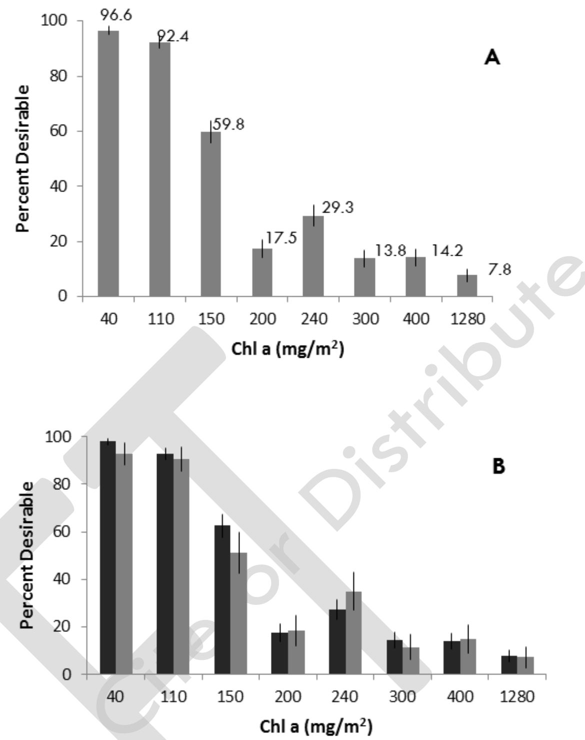


Figure ES3. A) Percent desirable benthic algae response from all Utah survey participants. B) Percent desirable benthic algae responses from users (black) and non- user (grey) groups showing similarity in responses. Error bars indicate 95% confidence interval.

Recreation Responses

Excessive nutrients can also degrade recreation uses. In some cases, nutrient-related degradation of recreation uses is caused by increases in pathogens or biological toxins. More frequently, particularly for small to moderate size streams, recreation uses are potentially degraded by decreased aesthetics from excess benthic algae. To quantify the latter, an investigation was conducted to relate aesthetics, recreation and algae growth (Chapter 7). Surveys were mailed to 2,700 randomly selected Utah households that included pictures of streams with different concentrations of benthic algae. For each randomly ordered picture, survey participants were asked whether they would consider the depicted conditions to be desirable or undesirable to their recreation experience. The majority of the 628 respondents indicated a shift from desirable to undesirable conditions as algae densities increased from 110 to 150 mg chl-a/m^2. These opinions did not differ between citizens who reported recreating on streams (users) vs. those who do not (non-users).

Application of Thresholds

The third section of the report is intended to provide the technical basis for the application of the thresholds derived in the technical basis to UDWQ regulatory programs. Currently, this section includes two specific applications: the development of headwater numeric criteria (Chapter 10) and considerations for the ongoing application of water quality indicators to the development of numeric criteria for streams lower in the watershed (Chapter 11). Another chapter that describes the use of process-based models for site-specific NNC is in development and will be added to this section when it is complete.

With respect to headwater NNC development, Chapter 10 explores two questions: 1) Is there a need for subclasses of streams to account for natural variation?, and 2) How do nutrient concentrations among headwater streams compare with those obtained from designated reference sites or from statewide estimates?. To explore the need for subclasses, watershed scale descriptors of natural environmental gradients were used to divide 89 headwater reference streams into two classes that were as physically distinct as possible. Neither P (as TP) nor N (as Dissolved Inorganic Nitrogen (DIN) significantly differed between these physically distinct subclasses, which allowed UDWQ to conclude that further classification was not necessary for headwater criteria development. Not surprisingly, average nutrient concentrations among headwater reference sites were low for TP = 0.017 (±0.013-0.022) and dissolved inorganic nitrogen (DIN) = 0.192 mg/L (±0.125-0.259). Nutrient concentrations were also low among all headwater streams sampled over the previous 10-years. The median TN for headwater streams was 0.21 mg/L (n=385), which was only slightly lower than estimates of the median of all streams statewide (0.25 mg/L). In contrast, the median headwater TP of 0.012 mg/L was appreciably lower than estimates for the median TP of 0.04 mg/L of streams statewide.

Regional stressor-response patterns provide insights that can inform management objectives, but they have limitations. In particular, these approaches are unable to capture the influence of covariates on stressor-response relationships, particularly those that vary across local, site-specific scales. Accordingly, UDWQ has opted to generate site-specific NNC for stream outside of headwaters. This decision provides an opportunity to address several sources of uncertainty in stressor-response relationships that are difficult, if not impossible to address at a regional scale. Specifically, site-specific investigations offer an opportunity to embrace the intrinsic complexity of streams, to reduce uncertainty and strengthen causal stressor-response inferences. These site-specific investigations also provide the opportunity to evaluate the relative influence of nutrients and other stressors on ecological responses. The potential power of these site-specific investigations requires a carefully crafted study design, which involves consideration of as many potential sources of variation as possible. Specific guidance with regard to these considerations is provided in Chapter 11.

In concordance with adaptive management principles, the work described in this report represents on step among many in the ongoing development of Utah's nutrient reduction strategy. The stressor-response relationships developed in this report will continue to be refined as site-specific investigations continue. New indicators will likely be developed. These improvements will continue to improve the accuracy of nutrient-related assessments and the defensibility of site-specific NNC. Practically speaking, these tasks—among others—are best achieved by embracing principles of collaborative management. Many have a direct interest in setting appropriately protective NNC. Our ability to reduce scientific uncertainty hinges upon translating these collective interests into a shared understanding of the nutrient-related problems and appropriate regulatory responses.

Table of Contents

SECTION 3: APPLICATION TO REGULATORY PROGRAMS

SECTION OVERVIEW AND BACKGROUND

AMBIENT NITROGEN AND PHOSPHORUS IN HEADWATER STREAMS

ACCOUNTING FOR COMPLEXITY, UNCERTAINTY, VARIABILITY, AND COVARIABLES IN SITE-SPECIFIC ANALYSES: THE PATH FORWARD

LIST OF FIGURES

Figure Description	**Page Number**
with low (<48.76 g AFDM/m[2]) and high organic matter.	
Figure 6.1. Significant indicator diatom taxa plotted in order of relative sensitivity to nutrients.	113
Figure 6.2. Significant indicator macroinvertebrate taxa plotted in order of their relative sensitivity to nutrients.	115
Figure 6.3. Box plots of numeric O/E scores for sites above and below nutrient thresholds.	117
Figure 6.4. Prediction probability curves generated from ROC analysis to evaluate the predictive accuracy of nutrient thresholds.	118
Figure 7.1. Percent of survey respondents who viewed various levels of benthic algae as desirable conditions for recreation.	127
Figure 7.2. Photographs of streams and associated chlorophyll concentrations that were used for recreation use surveys.	130
Figure 9.1. Thresholds of all nutrient indicators for TN and TP that were derived through the investigations described in this report.	132
Figure 9.2. Aggregated Western Mountains (WMT) and Xeric Omernik Ecoregions used in reference site analysis.	135
Figure10.1. Utah's Antidegradation category 1 and 2 boundaries that UDWQ is using to demark headwater streams,	151
Figure 10.2. A color coded map depicting the two groups of watersheds (12-digit HUCs) determined from k-means cluster.	154
Figure 10.3. Results of PCA analysis depicting the physical and environmental factors that best classify 12 digit HUCs into two distinct groups.	156
Figure 10.4. Boxplots showing distributions of total phosphorus (TP) and total inorganic nitrogen (DIN) between the two physically distinct groups of reference sites (from k-clustering).	159
Figure 10.5. Results of Random Forest models that predict background total nitrogen (TN) and phosphorus (TP) concentrations from observations made a reference sites under baseflow conditions.	162-164
Figure 11.1. Study site locations within the state of Utah.	192
Figure 11.2. Generalized data collection locations for the model development portion of the study	197
Figure 11.3. Logic used in estimating VSS and ISS from TSS, followed by logic for estimating detritus from VSS.	216

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

Acronym/ Abbreviation	Description
AFDM	Ash Free Dry Mass
ANOVA	Analysis of Variance
ATP	Adenosine triphosphate
AUC	Area Under the Curve
awch	available water holding capacity of soils
BACI	Before, After, Control, Impact study design
BMP	Best Management Practice
CBOM	Course Benthic Organic Matter
CDF	Cumulative Distribution Function
chl-a	Chlorophyll *a*
CI	Confidence Interval
DNA	Deoxyribonucleic acid
DO	Dissolved Oxygen
elev	Elevation at the bottom of the watershed
ER	Ecosystem Respiration
FBOM	Fine Benthic Organic Matter
GPP	Gross Primary Production
GRTS	Generalized Random Tessellation Stratified
GSL	Great Salt Lake
HUC	Hydrologic Unit Code, a cataloging code used to distinguish watersheds
IBI	Index of Biological Integrity
kfact	soil erodability factor
MEANP	Mean annual precipitation
MSE	Mean Square Error
N	Nitrogen, expressed in mg/l throughout this report
NDS	Nutrient Diffusing Substrates
NHST	Null Hypothesis Significance Testing
NMDS	Non-Metric Multidimentional Scaling
NNC	Numeric Nutrient Criteria
NTU	Nephelometric Turbidity Units
O/E	A biological assessment metric that compares the number of taxa observed at a site sans human disturbance, E, to those that were predicted to occur, O.
OM	Organic Matter
P	Phosphorus, expressed in mg/l throughout the report
PAR	Photosynthetically Active Radiation
PCA	Principal Components Analysis
perm	Water permeability of soils

Acronym/ Abbreviation	**Description**
POTW	Publically Owned Treatment Works
PRISM	Physiographically Sensitive Mapping of temperature and precipitation
RCC	River Continuum Concept
RNA	Ribonucleic acid
ROC	Receivers Operator Curve
rockdepth	depth of soil to bedrock
SAP	Sample Analysis Plan
SEM	Structural Equation Modeling
SOP	Standard Operating Procedure
S-R	Stressor-Response relationship, in this case nutrients-ecological responses
SRS	Stratified Random Sample
STATSGO	State Soil Geographic Database
STORET	STOrage and RETrieval water quality database maintained by USEPA
TIN	Total Inorganic Nitrogen
TITAN	Threshold Indicator Taxon Analysis
TMDL	Total Maximum Daily Load
TMEAN	annual mean predicted air temperature
TN	Total Nitrogen
TP	Total Phosphorus, expressed in mg/l throughout the report
UAC	Utah Administrative Code
UDWQ	Utah Division of Water Quality
USEPA	United States Environmental Protection Agency
USGS	United State Geologic Survey
WQS	Water Quality Standard

CHAPTER 1

BACKGROUND: AN INCREMENTAL APPROACH TOWARD NUMERIC NUTRIENT CRITERIA

Introduction

The United States Environmental Protection Agency (USEPA) and Utah's Division of Water Quality have made the reduction of nitrogen and phosphorus (nutrients) pollution to Utah's waters a priority. To this end, Utah's Division of Water Quality (UDWQ) launched a series of investigations aimed at developing methods for quantifying several potentially deleterious ecological responses to these nutrients. This chapter provides a broad background on many of the issues that underpin the need for the technical investigations described in this report. We start with a brief review of nutrient-related water quality problems that have raised worldwide concerns over nutrient pollution. Next, we review water quality standards, specifically numeric nutrient criteria (NNC), as one regulatory approach for addressing these concerns.

Addressing Excess Nutrients: A National Water Quality Priority

The USEPA has identified nutrient pollution as a national concern citing several water quality issues, including:

- About one-third of the nation's rivers and streams were impaired for either nitrogen (N) or phosphorus (P) (USEPA 2006) and half have moderate to high levels of excess nitrogen and phosphorus;
- Downstream, 78% of coastal waters exhibit effects of excess nutrients and the extent of "Dead zones"—large areas that lack dissolved oxygen —continue to increase in the Gulf Coast and elsewhere;
- In lakes and reservoirs, algae blooms are on the rise and their associated toxins have potentially serious human health and ecological effects;
- In groundwater, nitrate drinking water violations have doubled over the past eight years.

These concerns are reflected in legal challenges to USEPA's progress on this issue. The USEPA continues to face numerous legal challenges associated with nutrient water quality problems and the regulatory approaches employed to address the problem. In some cases these suits allege that USEPA has failed to meet its statutory obligations by not imposing stricter regulations to address the problem, whereas other cases argue that the policies that USEPA has instigated exceed their regulatory authority. The particulars of these cases are less important that the overarching message: nutrients are a critically important water quality concern that demands attention on the national stage.

While many states, including Utah, conduct water quality assessments based on indicators that can be used to infer nutrient-related ecological responses (e.g., dissolved oxygen, pH), USEPA's position has been that these approaches are not adequately protecting beneficial uses. Instead, USEPA believes that comprehensive nutrient reduction programs are necessary to protect aquatic ecosystems (Stoner 2011). This policy directs each State is to develop a nutrient reduction strategy that meets eight specific objectives that, in turn, seek to meet sever key goals, including: establishing watershed-specific nutrient reduction targets, addressing important nutrient sources, and making progress toward the development of numeric nutrient criteria (NNC) that set firm limits on nutrient concentrations for all aquatic ecosystems.

Addressing Excess Nutrients: A Water Quality Priority for Utah

Utah is the second driest state in the nation and protecting this important resource is critically important. Utah's Division of Water Quality (UDWQ) continues to identify reservoirs and streams with nutrient-related water quality problems. When these problems arise, they limit or prohibit the use of the water for drinking water, recreation or aquatic life uses. Yet, the root causes of these problems are human activities—principally stormwater from our cities, treated wastewater, atmospheric deposition, and runoff from some agricultural operations—that are key sources of nutrients, but also critically important to our quality of life and our economy.

IMPORTANT NUTRIENT SOURCES	
Natural	**Human-Caused**
• **Erosion from nutrient-bearing rocks and soils** • **Atmospheric deposition** • **Wildlife waste** • **Decomposition of plants and algae**	• Septic tank leachate • Stormwater • Runoff from pasture or range • Runoff from agricultural fields including irrigation return flow • Discharges of treated wastewater • Overflow from combined storm and sanitary sewers • Industrial discharges

The need to address these conflicting needs is particularly important because Utah's population is expected to double by 2050, which will require increased dependence on lakes and rivers as sources of culinary water. If we do not address these concerns now, we risk jeopardizing the quality of life for future generations. Protecting the quantity of water resources is important, but these efforts are of little value if the quality of water is insufficient to meet our needs. Clearly a comprehensive plan to address nutrient pollution needs to incorporate consideration of the costs and benefits to Utah's citizens. UDWQ believes that Utahns are best positioned to craft an effective nutrient reduction strategy that reflects Utah's unique waters.

Utah's Division of Water Quality and the Environmental Protection Agency have made addressing nitrogen and phosphorus pollution a water quality priority because addressing this growing water quality concern is needed to ensure ongoing protection of the designated uses assigned to our lakes and rivers and to ensure the quality of life for future generations.

NNC are among the most efficient regulatory tools available to address N and P pollution. From UDWQ's perspective, established NNC would make several water quality regulatory programs more efficient because they would provide clear objectives that could be incorporated into discharge permits and TMDL endpoints. From the perspective of the regulatory community, NNC would clearly define long-term objectives and allow facilities to more easily establish water quality trading programs. In practice, however, the NNC that have been recently been promulgated by other states are well below treatment thresholds that are pratcially achievable. This disconnect between these two important values—the needs of aquatic ecosystems vs. those of communities—is one example of the cause of much of the controversy surrounding efforts to address nutrient pollution in surface waters. On one hand, our Publically Owned Treatment Works (POTWs) provide a service that the World Health Organization considers the most important medical advancement of the previous century. Yet, on the other hand, protective concentrations of NNC often cannot be met or would be extremely costly to achieve. NNC and other related water quality programs seek a balance among these conflicting needs, but there is a general lack of stakeholder concensus among the relative importance of conflicting values.

For the past several years, UDWQ has been meeting with leaders representing diverse interests to draft a comprehensive nutrient reduction strategy (see www.nutrients.utah.gov). The plan is multifaceted and includes several programs that strive for equitable solutions to nutrient pollution problems. Among these programs is a call for NNC that are supported by a scientific understanding of ecological responses to nutrient enrichment specific to Utah. This requires development of new approaches that will allow UDWQ to better quantify potentially deleterious ecological responses to nutrient pollution in Utah's streams. In response to this need, UDWQ consulted with international experts, and determined that improved understanding require multiple ecological responses, because the effects of nutrients on designated uses is both varied and complex.

This report summarizes findings from a study designed, and subsequently funded by Utah's Water Quality Board, to establish water quality indicators that could be used to better assess the effects of nutrient enrichment in Utah streams. UDWQ evaluated a wide range of indicators and selected a subset for detailed evaluation that met the following criteria:

- ***reflect support aquatic life uses—are robust indicators of stream condition***
- ***are scientifically defensible—well established in the scientific literature***
- ***could be directly linked to nutrients through conceptual or mechanistic models***
- ***were practical—could be routinely collected by UDWQ staff***

After selecting and collecting appropriate ecological responses the next step is the development of stressor-response models that quantify the relationships among N and P stream concentrations and several ecological responses. The majority of this report (Sections 2 and 3) describes the development of these stressor-response relationships and associated N and P thresholds. These empirical stressor-response models are intended to meet several objectives of Utah's nutrient reduction strategy. First, they will help define N and P concentrations that are associated, on average, with deleterious ecological responses with the potential to impair aquatic life designated uses. Depending on the location of the stream, nutrient response thresholds will provide the technical rationale for numeric criteria in Utah's headwater streams, or as numeric indicators of potential water quality problems elsewhere. Second, the simultaneous interpretation of multiple responses provides UDWQ with insight into the specific processes that are causing nutrient-related problems. These insights, together with mechanistic models (Neilson et al. 2012), will inform the design of site-specific standard studies and will inform the selection of specific remediation practices that are most likely to resolve nutrient-related problems. Finally, we expand the scope of the study to consider coupled human-ecological relationships with a survey that aimed to quantify the extent to which recreation uses are potentially affected by the presence of nuisance algae (Chapter 7).

How do excess nutrients degrade streams?

There are numerous ways that anthropogenic eutrophication can potentially have deleterious impacts to aquatic life, recreation or drinking water uses. Several authors have reviewed the multiple pathways that link excess N and P to degradation of aquatic ecosystems (e.g., Dodds 2006, Smith et al. 1999, Yuan et al. 2010). These links are illustrated in the conceptual model in Figure 1.1. While not appropriate for all streams, nor all responses, the model is useful because it includes several important considerations, including: the importance of numerous sources, modifying factors (i.e., covariates), ecosystem responses, biological responses, and interacting stressors. While a literature review of nutrient effects on ecosystems is unnecessary here, we briefly summarize several investigations that most directly relate to Utah's approaches for development of Numeric Nutrient Criteria (NNC) and associated assessment tools.

One principle pathway through which excess nutrients can degrade stream ecosystems is via increases in primary production: increased growth of macrophytes or algae. In larger, soft-bedded rivers this growth is typically manifest either as macrophytes or phytoplankton (Vanote et al. 1980, Whiles and Dodds 2002), whereas a shift from diatoms to filamentous algae is often observed in cobble-bedded streams (Slavik et al. 2004). Consequently, investigators frequently use alterations of algal assemblage composition to document nutrient-related effects (see Chapter 6). Initially—as nutrients increase beyond background concentrations—an increase in primary production poses few problems and can even be beneficial to stream ecosystems. Eventually, however, excess growth can cause several problems, the severity of which depends greatly on the site-specific biological (Haapala et al. 2001, Biggs et al., 2000) and physicochemical characteristics (Biggs 2000, Townsend et al. 2012, Lohman et al. 1992).

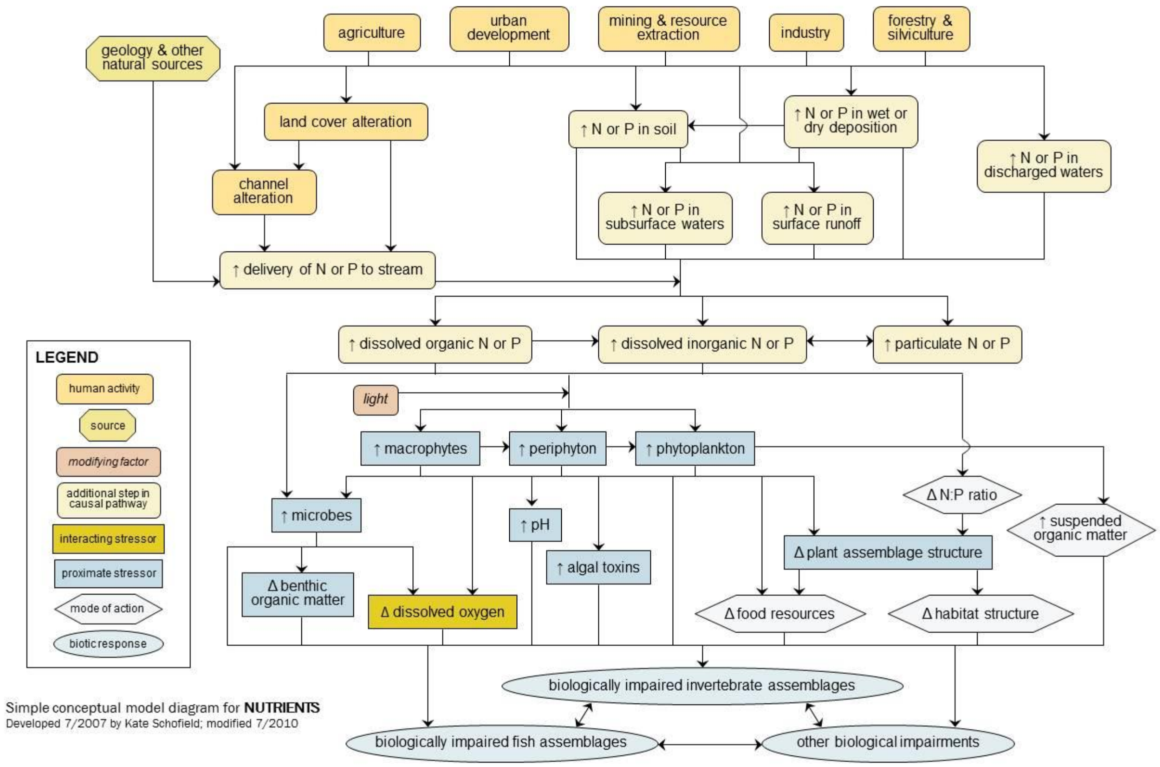


Figure 1.1. Conceptual model of linkages from nutrient sources to and beneficial uses (USEPA 2000).

One important ecological effect of eutrophication is reduced dissolved oxygen (DO) concentrations within streams. Increases in primary production correspondingly increase photosynthesis rates and daytime DO production. At night, all organisms continue to respire, which consumes DO. At night, DO consumption by aquatic biota is not offset by the DO produced by photosynthesis, so water column DO concentration declines. If atmospheric reaeration cannot compensate for these DO losses, then DO concentration within the stream can fall to levels that are unhealthy, even deadly, to some stream biota. Eutrophication, caused by excess N and P, exacerbates low nighttime DO because heterotrophic bacterial and fungal productivity can also be stimulated by nutrient enrichment (Tank and Winterbourn 1996; Chapter 4). Also, increases in plant or algae growth that sometimes occurs in high nutrient environments creates additional carbon, which ultimately leads to greater secondary production and an associated increase in respiration (Chapter 6). Particularly in temperate streams, these carbon increases have the potential to create low DO conditions during autumn senescence of algae or macrophytes, which provides a large pulse of labile carbon (Suplee et al. 2012). Overall, the close coupling of DO to both production and respiration provides a relatively simple way to obtain reach-scale estimates of both processes (Chapter 5).

Direct acute or chronic effects of low DO to stream biota are not the only deleterious impacts to stream ecosystems. Low DO decreases the sequestration rates of some heavy metals to sediments which creates additional threats to stream biota, especially considering that they are already under stress. Similarly, low DO sometimes interacts with other stressors (i.e., ammonia), which results in an increased threat to stream biota.

Another consequence of excessive primary production (Chapter 4) is degradation of stream habitat. For instance, a high abundance of filamentous algae alters benthic flow characteristics and traps fine sediment (Slavik et al. 2004). Eventually this fine sediment fills benthic interstitial spaces, which is critical habitat for macroinvertebrates that are the base energy source of stream fishes (Wallace and Webster 1996). Special physiological adaptations are required to consume filamentous algae, or the organic matter trapped within, which alters the abundance and distribution of macroinvertebrate taxa (Meritt et al. 2008, Dudley et al. 2000). Moreover, the quality of algae food resources, as expressed by C:P or C:N, often declines in high nutrient environments (King et al. 2000). In clear, soft-bedded rivers nutrient enrichment can lead to increased macrophyte growth, which can improve habitat conditions (Fritz et al. 2004), provided that these changes do not lead to low levels of dissolved oxygen (see below). Although others have noted declines in macrophyte cover at streams with very high nutrient concentrations (King et al. 2000). Increased phytoplankton growth decreases water clarity, which can cause a shift in the base of fish food webs from benthic to pelagic sources (Vanote et al. 1980). Such habitat changes alter the diversity and abundance of macroinvertebrates, which provides an indirect metric of the net effects of these habitat alterations, and other sources of stress (Chapter 6). Excess fine sediment can also directly affect fish through a reduction in the availability of spawning gravel, although filamentous algae is not entirely detrimental to fish because it also provides valuable fish cover. One result of these changes is alterations to the types and relative abundance of organic matter stored in the stream (Chapter 5). Most people find streams with excess production to be less aesthetically appealing, which degrades recreation uses (Suplee et al. 2012; Chapter 7).

Eventually, incremental increases in nutrients cannot further increase primary production because other factors become more important in limiting growth of plants and algae (Bernot and Dodds 2005). Similarly, an increase in either N or P can potentially switch the nutrient that most limits production (Dodds et al. 1997, Chapter 3). In streams, saturation of ***both*** N and P rarely occurs naturally, so saturation points represent conditions where nutrients are almost certainly contributing to the degradation of stream ecosystems (Chapter 3). The saturation concentration of N and P also has important restoration considerations because future nutrient reductions would be unlikely to improve stream conditions until they result in ambient concentrations below saturation thresholds. Nutrients not incorporated into biota are transported downstream, which sometimes cause problems far from nutrient sources. Nutrients can also accumulate locally, typically in stream sediment. High concentrations of N or P within stream sediment are important to riverine management, because internal loading of nutrient from sediment to the water column could offset external point or non-point source nutrient reductions resulting in a lag-time in stream response.

The extent to which excessive nutrients result in any of these problems depends greatly on site-specific characteristics. Natural conditions (e.g., channel shading, water temperature) mediate among stream differences in nutrient responses. For instance, higher gradient streams are buffered against potential DO problems because they typically have higher canopy cover, which lowers primary production, and increases atmospheric reaeration. Higher gradient streams are also less likely to accumulate N, P or C, which afford them natural protection from chronic nutrient inputs. Structural responses are also affected by these natural environmental gradients. For instance, among stream differences in temperature or slope directly alters the distribution of macroinvertebrate or algae taxa.

Other site-specific conditions that modify nutrient responses are human-caused. For instance, people sometimes remove riparian vegetation, which can increase primary production rates. Other landscape-level changes, particularly urbanization, alter the transport and storage of nutrients in streams (Paul and Myer, 2001). These land uses also cause other sources of stress that can have both synergistic and antagonistic effects to stream structure and function. When feasible, nutrient restoration plans should consider and incorporate improvements to all of the degraded physicochemical characteristics that operate in concert with nutrients to degrade stream conditions. However, in some cases such changes are practically irreversible, in which case new goals—water column concentrations or ecological response thresholds—will need to be established.

Numeric Nutrient Criteria: One Regulatory Option for Addressing Nutrient Pollution

As previously mentioned, the studies in this report are intended to inform the incorporation of nutrients into several regulatory programs that UDWQ already uses to address water pollution concerns. This section focuses on water quality standards (WQSs) because the development of NNC is the ultimate goal of these and follow-up investigations.

Water Quality Standards

The Clean Water Act (CWA) authorizes States and Tribes to develop and adopt WQSs to protect the chemical, physical and biological integrity of their surface waters. WQSs consist of beneficial uses of a waterbody, water quality criteria to protect those beneficial uses, and antidegradation policies that aim to protect high quality waters. States establish designated uses for waterbodies, including such uses as recreation, agriculture, aquatic wildlife, etc. Water quality criteria—both narrative and numeric— define specific conditions that must be maintained to protect these uses. Narrative criteria describe conditions that should be avoided—or conversely those that must be maintained— in order for a water body's uses to remain protected. In contrast, numeric criteria (UAC R317-2-14) provide specific pollutant concentrations (e.g., magnitude) that must not be exceeded. Narrative criteria are intentionally broad, which has the advantage of capturing deleterious water quality conditions that are difficult to quantify. As a result, Utah's narrative criterion already addresses several deleterious nutrient effects, with the inclusion of statements like "scum", "undesirable aquatic life", "objectionable taste or odor"; yet, Utah's Division of Water Quality (UDWQ) has found it difficult to develop effective regulations exclusively from such subjective goals. For instance, it is not always clear how to translate narrative statements to permit limits. Narrative criteria are also problematic to the regulated community because they complicate business planning in the form of uncertainty. Unfortunately, the net result of ambiguous regulations is often legal challenges. If nutrient related water quality problems are to be rectified and prevented, Utah needs a path that will ultimately lead to numeric N and P criteria.

Utah's Narrative Criteria (UAC R317-2-7.3)

It shall be unlawful, and a violation of these regulations, for any person to discharge or place any waste or other substance in such a way as will be or may become offensive such as unnatural deposits, floating debris, oil, scum or other nuisances such as color, odor or taste; or cause conditions which produce undesirable aquatic life or which produce objectionable tastes in edible aquatic organisms; or result in concentrations or combinations of substances which produce undesirable physiological responses in desirable resident fish, or other desirable aquatic life, or undesirable human health effects, as determined by bioassay or other tests performed in accordance with standard procedures.

To date, most numeric water quality criteria aim to protect designated uses from toxic substances. The analyses that underpin these criteria are a compilation and summation of toxicological laboratory studies. Because these criteria are derived from laboratory work, the toxicity of many of these compounds is not place dependent (although some have modifiers). In addition, these criteria are broadly toxic to many different organisms, so criteria for these substances are amenable to broadly applicable national recommendations from USEPA. Nutrients are different. While some forms of nitrogen (i.e., ammonia, nitrate) are toxic to humans and aquatic biota, excess nutrients—particularly the macronutrients nitrogen and phosphorus—can degrade a stream's designated uses at concentrations much lower than toxic levels (Table 1.1). Standards to protect against the many non-toxic deleterious effects of nutrients are an important part of protecting aquatic life and recreation uses, but unlike those developed for toxic compounds these standards cannot be derived from traditional laboratory toxicology test results. As a result, alternative approaches to setting NNC are required.

HOW ARE NUMERIC NUTRIENT CRITERIA DEVELOPED?

The USEPA recommends three approaches for setting scientifically defensible NNC: 1) distributions which define NNC based on the distribution of N and P concentrations among reference sites; 2) mechanistic modeling, which use various models to estimate the N and P that are predicted to exceed other predefined water quality indicators or criteria (i.e., DO, pH, benthic algae cover); and 3) stressor-response analysis (USEPA 2000). Each approach has strengths and weaknesses. Moreover, the approaches are not mutually exclusive. In many cases, results from one method can be used to inform the results obtained from others. As a result, UDWQ has explored each method and plans context-dependent approaches for their iterative integration into water quality programs.

Distributional approaches are one common way that NNC are developed (for details and applications see Herlihy et al. 2008, Paulsen et al. 2008 and Hawkins et al. 2010). This approach uses data obtained from numerous reference sites—streams that have experienced minimum anthropogenic disturbance —within a predefined region wherein expected conditions are expected to be roughly uniform—typically ecoregions. Benchmarks are subsequently derived from these reference data with a statistical evaluation that considers natural variation among all sites. The values used to demark classes are somewhat arbitrary, but USEPA methods generally prescribe benchmarks for NNC as the 75th percentile of all concentrations observed among all reference sites. One implicit assumption of this method is that some reference sites are either actually degraded (in which case they would, by definition, not be in reference condition), or otherwise are atypically high in nutrients due to atypical natural conditions. The latter case could be remedied with site-specific NNC, but this diverts resources from fixing water quality problems. Another problem is that distributional approaches do not articulate how nutrients above these thresholds alter designated uses. Nevertheless, such approaches do allow managers to easily establish N or P benchmarks for a region of interest, which subsequently provides insight into background conditions.

Process-based water quality models, sometimes called mechanistic models, are another approach to setting NNC. These models use predefined mathematical relationships to couple chemical, physical and biological processes (see Chapra 1997, Edinger 2002 for reviews). To create these models, one selects the appropriate water quality goals based on responses of greatest interest, and then calibrates the model with monitoring data or literature values to constrain important factors within the model algorithms. NNC can then be "backed out" of the models by asking what concentrations of N or P cause modeled violations in the water quality goal. There are many water quality models that are used to generate NNC. The Water Environment Federation recent conducted a review and found 30 potential models that could be used for these purposes (WEF 2013, initiative Link1T11). These models differ in complexity, each incorporating and subsequently emphasizing the relative importance of different biogeochemical processes. These models can be interpreted in mechanistic contexts, which can help interpret stressor-response relationships (SAB 2010). Another advantage of these models is that they always generate an endpoint, in this context a concentration of N or P that can be used to set NNC. Of course this simplicity can also be a disadvantage because the models vary in their treatment of uncertainty and variability. In general, models, with their intrinsic focus on processes, tend to most accurately predict physical conditions (i.e., temperature, DO) followed by water chemistry (i.e., nutrients) and then biological endpoints. The mathematician George Box (1987) famously captured the interpretability vs. accuracy tradeoffs when he stated, "Essentially, all models are wrong, but some are useful." To date, process-based models have not been widely applied for setting NNC, especially regional NNC, largely because they need to be calibrated on a site-specific basis. As a result, UDWQ has primarily used these water quality models for setting TMDL endpoints or permit limits. Another disadvantage of models is that input and calibration data requirements are typically high, which makes them resource-intensive. Nevertheless, UDWQ considers models to be valuable tools and we intend to apply them within the context of follow-up site-specific investigations.

Stressor-response methods are currently the most frequently used methods to develop regional NNC. These empirical methods relate stressors (e.g., N or P) to ecological responses such as changes in biological composition (ecosystem structure) or biogeochemical process (ecosystem functions). To establish these relationships, stressor and response data are obtained from numerous sites that together encompass the range of conditions within a region of interest. Next, statistical models are used to establish thresholds that define specific stressor conditions (i.e., concentrations of N and P) that are most strongly associated with changes in ecological responses. In essence, these methods seek stressor values that best maximize among-group differences in ecological responses, while also maximizing within-group homogeneity. Once thresholds have been established, they can be related to protection of aquatic life uses through existing numeric criteria (i.e., pH, DO), biological assessment outputs, or with conceptual models that describe direct and indirect connections to designated uses (Table 1.1). The 2nd section of this report (Chapters 2 – 5) describes stressor-response models related to ecosystem functions, whereas models that relate N and P to measures of ecosystem structure are described in the 3rd Section (Chapter 6).

TECHNICAL CHALLENGES WITH ESTABLISHING NUMERIC NUTRIENT CRITERIA

Although the USEPA has emphasized the importance of States developing NNC for over a decade, most States have yet to adopt NNC for all waterbody types (see see USEPA's N and P policy online materials for details). The reasons for the lack of universal adoption of numeric criteria are many, and include both technical challenges associated with the derivation of numeric criteria and the sociopolitical ramifications surrounding their implementation. This section focuses on the former. Clearly, national "one-size fits all" NNCs for the Nation are inappropriate, which means that the States are shouldering much of the burden to overcome these technical and political challenges.

Nutrients are Natural and Important Components of Stream Ecosystems

One technical challenge with NNC development is that while excess N and P can cause degradation of aquatic ecosystems, N and P are also naturally occurring and essential macronutrients. All life depends upon inorganic phosphorus because, among other reasons, both elements are structural components of DNA and RNA. In addition, cellular processes that require energy almost always use adenosine triphosphate (ATP) to transport energy in the form of chemical bonds. Phospholipids, a major component of cellular membranes, also contain phosphorous. Unlike N, the phosphorus cycle does not contain a major atmospheric component; instead, it is stored mainly in rocks and sediments. Phosphorus becomes available primarily through weathering and is therefore often naturally rare in many ecosystems, although exceptions exist within Utah watersheds where the lithology contains rocks with high P concentrations. Elsewhere, P in mined from soils that are naturally enriched soils and sold as fertilizer. Runoff from excessive application of these fertilizers is one reason why phosphorus continues to increase in aquatic ecosystems worldwide. Nitrogen is also present in DNA and RNA as it is an essential structural component of nucleic acids and amino acids (hence proteins). Although nitrogen gas comprises the majority of the earth's atmosphere (~75%) it is unavailable to biota until it is "fixed" by specialized microorganisms, or by people via Haber-Bosch fertilizer production. Because of this, the bioavailable forms of nitrogen are often at relatively low concentrations in unperturbed aquatic ecosystems. Together, these factors complicate efforts to identify nutrient concentrations that are natural and necessary for healthy stream ecosystems, or conversely concentrations that cause impairments to beneficial uses.

Another reason States have yet to universally adopt NNC criteria is that background (natural) N and P concentrations vary greatly from region-to-region, as does the magnitude of potentially deleterious ecological responses. For example, the USEPA, in collaboration with States and Tribes, conducted the first statistically valid sampling design of the Nation's streams using reference streams (minimal anthropogenic disturbance) to set benchmarks for numerous stressors (USEPA 2006). In this study the United States was divided into nine aggregated ecoregions and the 95th percentile of reference condition streams (natural background) was used as a benchmark for "poor" condition. Total nitrogen 95th percentile concentrations ranged from 3.21 mg/L in the Temperate Plains ecoregion to 0.229 mg/L in the Western Mountains ecoregion. Total phosphorus 95th percentile concentrations varied from 0.338 to 0.036 mg/L among reference sites in the same ecoregions (Van Sickle and Paulsen 2008).

Regional Stressor-Response Relationships cannot Always Account for Covariates

Broad classifications, such as ecoregions, account for among stream differences in landscape-scale characteristics such as soils and precipitation. However, these classes may still be too broad to adequately capture important covariates that vary over much smaller spatial scales. After all, for at least several decades, stream ecologists have acknowledged that natural gradients in geomorphology such as stream size, width:depth ratios, slope and sinuosity naturally affect the structural and functional characteristics of stream ecosystems (i.e., Vanote et al. 1985, Rosgen 1994). Finer scale classifications could potentially avoid covariate bias, because it involves creating groups of streams with similar morphological characteristics. However, here again, each ecosystem response will likely respond differentially to environmental gradients. At some point, regional stressor-response generalizations need to be verified, and where appropriate modified, on a site-specific basis.

Water Chemistry needs to be interpreted in Concert with Ecological Responses

Not only are water column nutrients naturally variable, paradoxically they also are not necessarily reflective of the extent of nutrient enrichment. Especially during the growing season, sites with high primary production may actually have low concentrations of N and P in the water column because nutrients are incorporated into plants and algae. Water quality programs often collect total and dissolved N and P concentrations expressly to address this problem, however this approach cannot compensate for circumstances where the majority of primary production is benthic. Moreover, as previously discussed, the ecological consequences of excess nutrients also vary. There are numerous causal pathways whereby excessive nutrients can degrade stream ecosystems (Yuan et al. 2010) and the relative importance of each pathway varies spatially and temporally (Smith et al. 2006). Moreover, N- and P-driven transitions from "good" to "bad" ecological conditions are often gradual, which complicates demarcation of a single protective regional criterion, even within relatively homogeneous regions.

Most Degraded Streams are in Multiple-Stressor Environments

Another challenge is that nutrient pollution rarely occurs in the absence of other human-caused stressors to stream ecosystems. More commonly streams become degraded in response to multiple stressors, which are often difficult to decouple. In some circumstances, one stressor can exacerbate another. For instance, a denuded riparian corridor can exacerbate problems of excess primary production, which means that lower levels of N and P could degrade biological uses. Similarly, other stressors could mask the effects of excess N and P. For instance, excess turbidity caused by habitat degradation or stormwater can reduce light penetration and limit algae growth. Such circumstances do not necessarily preclude the development of nutrient stressor-response relationships, but they do complicate efforts to ascribe a specific cause. In most circumstances, distinguishing the relative effects of different stressors can only be accomplished on a site-specific basis, which UDWQ has incorporated into NNC program development. Nevertheless, even under multiple stressor conditions, regional stressor-response relationships, such as those described in this report, still provide useful measures of ecosystem health, particularly if several lines of evidence are considered concurrently.

Utah's Approach for the Development of NNC

As previously mentioned, NNC are a critical component of Utah's nutrient reduction strategy and ultimately UDWQ intends to establish appropriate NNC for all surface waters. However, NNC in isolation of a comprehensive nutrient reduction strategy will never address all of the problems associated with N and P water pollution. Water quality improvements also require a reasoned strategy for the *implementation* of NNC, including the reduction of N and P from both point and non-point sources. Implementation considerations must also be commensurate with the social and scientific complexities intrinsic within N and P water quality concerns. UDWQ has several ongoing efforts to address implementation complexities in the context of a comprehensive nutrient reduction strategy. A detailed explanation of implementation details is beyond the scope of this document, although the technical basis for several implementation efforts is provided in the final section of the report. Here we focus on UDWQ's plans for phased implementation of NNC, which is one component of Utah's adaptive approach to implementation of a nutrient reduction strategy.

Prioritization: A Phased Implementation of NNC for Utah's Waterbodies

FIRST STREAMS, THEN LAKES AND RESERVOIRS

UDWQ has initially focused on the refinement of N and P indicators and NNC development for stream ecosystems. Lakes and reservoirs are equally important, but in many respects UDWQ is already addressing nutrient problems in these waters. Utah's lake assessment methods evaluate several nutrient response metrics—such as the Trophic State Index (TSI), violations of numeric DO criteria and cyanobacteria dominance—to make determination about whether nutrients are degrading these waters. When eutrophic lakes are identified UDWQ develops Total Maximum Daily Loads (TMDLs) that specify in-lake nutrient concentrations and required nutrient load reductions for all sources. Another reason to focus on streams is that effective nutrient management in streams and rivers also affords some protection for lakes and reservoirs because streams are the primary external nutrient inputs to lentic ecosystems. However, in many cases, lakes may actually be more sensitive to excess nutrients than streams, which would require refining upstream NNC through TMDLs. Utah's approaches to address N and P pollution in Utah's lakes and reservoirs do need to be refined, and NNC ultimately established, but due to the considerations above UDWQ identified approaches for streams as an immediate need.

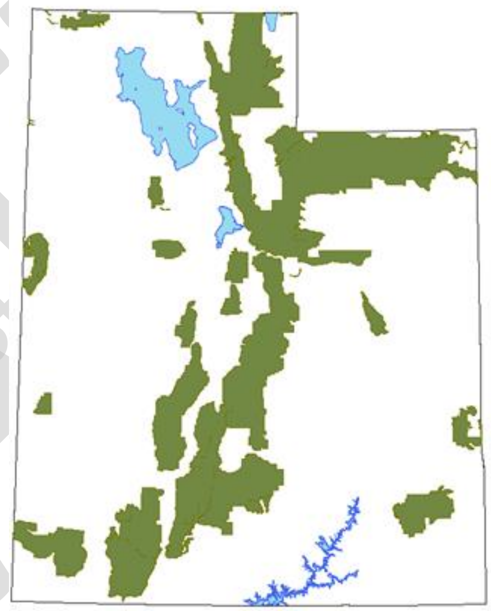


Figure 1.2. This map depicts Utah's antidegradation category 1 and 2 boundaries in green. UDWQ is proposing regional N and P numeric criteria for these waters as an early step in NNC development for waters of the state.

GREAT SALT LAKE: A UNIQUE ECOSYSTEM THAT REQUIRES SPECIAL CONSIDERATION

Additionally, specific nutrient requirements pertaining to Great Salt Lake must be addressed separately from other lakes because Great Salt Lake (GSL) is a unique hypersaline environment in which nutrient requirements and nutrient cycling are very different from freshwater lakes and typical marine environments. Simply put, the deleterious effects of excess nutrients on aquatic life uses that are observed in freshwater lakes may not be applicable to GSL because many sensitive organisms (i.e. fish) do not reside there. Moreover, a nutrient management strategy for freshwater lakes applied blindly to GSL could adversely effect the biota of the GSL and the essential ecosystem functions of the lake. Keystone species such as the brine shrimp (*Artemia franciscana)*, identified as a "Species of Protected Aquatic Wildlife" (Administrative Rule R657-52-1 and Rule R657-52-11), are fundamentally essential to the ecosystem of GSL. The GSL is a body of water whose ecosystem functions have hemispheric consequences. Because of the unique ecosystem characteristics, the hemispheric importance and protected keystone species of the GSL a rigorously detailed and tailored approach to addressing nutrient issues of the GSL will be implemented.

PRIORITIZATION OF HEADWATER STREAMS FOR NNC DEVELOPMENT

Initially, UDWQ will propose NNC for headwater streams. Elsewhere, UDWQ will promulgate NNC on a site-specific basis. In either case, the indicators described in this report will become integral to all nutrient reduction efforts because they form the basis for nutrient-specific monitoring and assessment approaches, which will allow UDWQ to identify and prioritize sites for follow-up site-specific investigations.

The decision to prioritize headwaters has a socioeconomic and technical basis. Utah's antidegradation rules identify areas of the State that contain high-quality headwater streams—often important drinking water sources —and has afforded them greater protection (Figure 1.2). In these waterbodies the State does not allow any discharges (Category 1), or in the rare cases where discharges are permitted they cannot exceed background conditions of the receiving water (Category 2). These waters occur mostly on United States Forest Service (USFS) lands, which have fewer anthropogenic stressors. As common recreation sites and drinking water sources, Utah's headwater streams have great economic importance. Technically, UDWQ is most confident that stressor-response thresholds are applicable to headwaters. Most of Utah's most pristine streams—and many reference sites—are located in headwaters. Most anthropogenic stressors occur lower in our watersheds. Stressor-response analyses need to encompass the range of nutrients within Utah, but ensuring that we included sites with the lowest and highest nutrient concentrations may have created systematic spatial bias. Moreover, the numeric thresholds established from these analyses are likely achievable for headwaters, whereas they may require adjustment in valley streams due to irreversible geomorphic or hydrologic modifications. Confirmation of indicators downstream of headwaters also requires an evaluation of covariates to confirm that deleterious structural and responses are the result of nutrient enrichment. Important covariates to examine include characteristics such as temperature, gradient, substrate size, riparian condition, and other human-caused stressors.

Development of Nutrient-Related Water Quality Indicators

As previously discussed, many physical and chemical stream characteristics interact to determine the extent to which nutrient inputs result in degradation of aquatic life, recreation or drinking water uses. In support of NNC development, UDWQ undertook a study to evaluate the potential for several ecosystem processes, or functions, to be combined with structural responses to more accurately characterize the effects of nutrients in Utah streams (Chapter 2). The functional responses selected for this investigation were not intended to be exhaustive. Instead, the focus was on ecological responses that are known to reflect nutrient enrichment and that UDWQ could most easily incorporate into ongoing monitoring and assessment programs. The framework also provides UDWQ with the ability to include other, potentially more complicated or costly responses into site-specific investigations.

UDWQ intends for the functional and structural response thresholds described in this report to serve two purposes. First, they will help define N and P concentrations (thresholds) that are associated, on average, with significant ecosystem responses. Numeric thresholds derived from these relationships will then provide the technical rationale for either numeric criteria or indicators of nutrient-related impairments. Second, the simultaneous interpretation of multiple responses provides insight into the specific processes that are causing nutrient-related effects. These insights, together with mechanistic models (Neilson et al. 2012), will inform the design of site-specific standard studies and will inform the selection of specific remediation practices that are most likely to resolve nutrient-related problems.

APPLICATION OF STRESSOR-RESPONSE MEASURES TO WATER QUALITY ASESSMENTS

One important outcome of the stressor-response linkages described in this report is the creation of a suite of tools that UDWQ can use to more accurately and completely identify streams with nutrient-related problems and derive appropriate site-specific NNC. We decided to explore multiple responses to provide UDWQ with a more holistic picture of the varied and often complex responses to cultural eutrophication. Our rationale was that multiple lines of evidence help UDWQ to more defensibly relate increases in nutrients to degradation of uses (Table 1.1).

Although many of these responses could be considered independently, UDWQ considers simultaneously and in context of defensible conceptual models supported by multiple lines of evidence. In general, this weight of evidence approach also addresses concerns about site-specific covariates that could potentially invalidate regional characteristics. Multiple responses collectively provide insights into linkages among nutrients, processes, and the degradation of designated uses. These evidence-based assessments will also elucidate areas of uncertainty that should be explored to develop effective approaches for reestablishing the biological integrity of degraded streams.

Practically speaking, the application of these multiple lines of evidence assessments is potentially resource intensive. As a result, UDWQ has developed tiered monitoring approaches that use routine monitoring information to screen sites for follow-up investigations. In this cases, when sites with high N or P are identified UDWQ will conduct additional monitoring to provide detailed measures structural and functional responses to eutrophication. If these data identify a nutrient-related impairment, then UDWQ will use these response indicators to inform the design of site-specific follow-up investigations for the purpose of establishing NNC and appropriate water quality targets that are direct measures of the conditions that need to be restored. Details of UDWQ's these monitoring methods are provided in *Utah's Strategic Monitoring Plan*, whereas detailed assessment methods will be described in the methods of the *Integrated Report*.

Report Organization

This report provides the technical basis for several elements of Utah's nutrient reduction strategy. While the report describes a suite of cohesive indicators, its constituent chapters were written to "stand alone", so that the indicators within each chapter can be easily incorporated into future sampling and analysis plans (SAPs) for site-specific investigations. We also wanted to make the report modular so that we could expand on this information as the nutrient program develops. The report is divided into three broad sections. The ***first section*** of the report (Chapters 2-5) describes the relationships among nutrients and responses that capture several important ecosystem functions. ***The second section***, (Chapters 6 and 7), describes relationships among nutrients and aquatic life (structural responses), then recreation uses, then provides summarizes the information gleaned from these stressor-response relationships (Chapter 8). The ***third section*** of the report provides the technical basis for the application of the nutrient indicators toward UDWQ's water quality programs.

Table 1.1 Links among indicators, nutrients, and Utah's Water Quality Standards (UAC R-317-2).

Indicator	**Description**	**Relationship to Nutrients**	**Tie to Standards**
Nutrient Diffusing Substrates (Chapter 3)	Response of benthic algal production to experimental N and P additions.	***Direct***: Algal growth is directly coupled to N and P.	***Narrative***: excess benthic algae can create "objectionable conditions" (Chapter 8) or conditions detrimental to desirable fish, but only with large accumulation.
Metabolism: Gross Primary Production (Chapter 4)	Reach-scale measure of the total primary production.	***Direct***: Plant and algae growth is directly coupled to N and P.	***Numeric***: Excess production can cause pH to exceed numeric criteria. Also, this metric directly relates to requirement for total dissolved gasses not to exceed 110% saturation, but not evaluated in the report. ***Narrative***: Can related to prohibited conditions such as undesirable tastes and odors, surface debris, and undesirable aquatic life.
Metabolism: Ecosystem respiration (Chapter 4)	Reach-scale measure of the respiration of all plants and animals.	***Direct***: Growth of heterotrophic microbes and fungi is directly coupled to nutrients.	***Numeric***: Directly relates to minimum Dissolved Oxygen (DO) requirements.
Organic Matter (Chapter 5)	Measures of various standing stocks (storage) of different types of organic matter.	***Both***: Direct if autochthonous (produced within the stream), indirect if allochthonous (from outside the stream).	***Numeric***: Directly relates to minimum Dissolved Oxygen (DO) requirements.
Alteration of the composition of macroinvertebrate and diatom assemblages (Chapter 6)	Differential responses of sensitive and tolerant taxa to increasing nutrients.	***Indirect***: Assemblages changes occur <u>following</u> nutrient-mediated modifications to physical and chemical habitat.	***Narrative***: Linked to biological assessments, which directly quantify biological integrity, a fundamental Clean Water Act goal.
Biological Assessments(O/E) (Chapter 6)	Estimates lost macroinvertebrate diversity— as the ratio taxa observed (O) to those expected (E) *sans* human disturbance.	***Indirect***: Assemblage changes occur <u>following</u> nutrient-mediated modifications to physical and chemical habitat.	**Numeric**: Designated use descriptions require protection of the food web. ***Narrative***: References to undesirable or nuisance organisms.
Aesthetics: Filamentous Algae (Chapter 7)	Survey of Utahns to evaluate the extent to which large algal blooms make river recreation less desirable.	***Direct***: Algae growth is directly couple to N & P. Filamentous algae become dominant in high nutrient, cobble-bedded stream.	***Numeric (proposed)***: Proposal to protect recreational uses with benthic chlorophyll a concentrations. ***Narrative***: References to undesirable or nuisance organisms.

Ecological Functions as Indicators of Nutrient Enrichment

SECTION 1

CHAPTER 2

FUNCTIONAL RESPONSES: STUDY DESIGN

Introduction

To date, most States and USEPA have either derived numeric nutrient criteria (NNC) from empirical stressor-response methods, frequency distribution methods, or mechanistic models. Of these, the frequency distribution method is the simplest because NNC are based on predefined percentiles—typically the 75^{th} percentile of reference sites or the 25^{th} of all sites of observed nitrogen (N) or phosphorus (P) within a region of interest (USEPA 2000). While some States have derived NNC with frequency distribution methods, most take umbrage with the fact that these approaches do not directly consider the condition of aquatic life designated uses. In contrast, stressor-response empirical models derive NNC from statistical thresholds that are based on linkages between nutrient concentrations and measures of ecological response that are direct or indirect measures of aquatic life uses. To our knowledge, States have almost exclusively evaluated structural ecological responses based upon alterations to the composition and relative abundance of aquatic assemblages (Weigel and Robertson 2008, MN PCA 2008, ME DEP 2009, FL DEP 1012). Although in some cases, mechanistic modeling approaches have been used to derive site-specific criteria or otherwise confirm regional stressor-response relationships (Flynn and Suplee 2011). Several aquatic assemblages that are known to be sensitive to nutrient enrichment have been evaluated, including macroinvertebrates (i.e., King and Richardson 2003), algae (i.e., Dodds et al. 2002), and less commonly fish (i.e., Wang et al. 2007). The focus on structural responses for NNC development reflects the long-standing use of these indicators in biological assessment programs by both State and Federal agencies, including UDWQ (see UDWQ 2010, 2014). Generally speaking, structural indicators are among the best indicators of overall health of aquatic ecosystems because they directly quantify key aspects of biological integrity (Karr 1981, Barbour et al. 2000). However, they create challenges to water resource managers under circumstances of multiple interacting stressors, because it is difficult to determine specific causes of degradation.

The study investigated two classes of ecological responses:
Functional Indicators: Measures of processes (or properties) that focus on physical-biological ecological linkages, in particular those that describe the flow of energy through ecosystems.

Structural Indicators: Measures of the composition and relative abundance of individuals in aquatic assemblages such as macroinvertebrates and fish.

Despite their strength in biological assessment programs, measure of structural condition—particularly those based on mancroinvertebrates and fish—may not be the best way to quantify ecological responses to anthropogenic eutrophication. Except under extreme conditions (i.e., ammonia toxicity), nutrients are usually not directly toxic to aquatic life; instead, there are intermediary effects of increased nutrients to stream process—or functions —that cause alterations to the structure of stream biota. Hence, measures of intermediary functional responses provide more direct, and potentially more accurate, measures of the impacts of nutrient enrichment. Indeed, several researchers have proposed augmenting biological assessments with direct measures of ecosystem processes (Grace and Imberger 2006). However, States, including UDWQ, have not traditionally incorporated such measures into routine stream monitoring programs due to resource constraints.

Over the past several years UDWQ has investigated the potential for several functional indicators to represent the effects of nutrient enrichment on stream ecosystems. Specifically, the study evaluated relationships among streams with varying N and P concentrations with three functional responses: nutrient limitation and saturation, algal production, stream metabolism and organic matter standing stocks. Initially, the principal study objective was to support development of NNC. As the study progressed, it was increasingly clear that for many streams numeric criteria would need to be derived on a site-specific basis because traditional classification methods were unable to sufficiently minimize natural variation in ecosystem responses at a regional scale (Chapter 12). Under circumstances where direct application of regional thresholds to specific sites is not appropriate, UDWQ will use the field and analytical methods developed for this study to augment ongoing monitoring and assessment programs to more accurately identify streams with nutrient-related water quality problems.

This chapter broadly describes the general study design and general methods that are broadly applicable across indicators. Specifically, we provide an overview of our study design and the collection of the underlying water chemistry data. More detailed methods specific are provided in the chapters of this report specific to each indicator and in the Standard Operating Procedures (SOPs) in the report appendices.

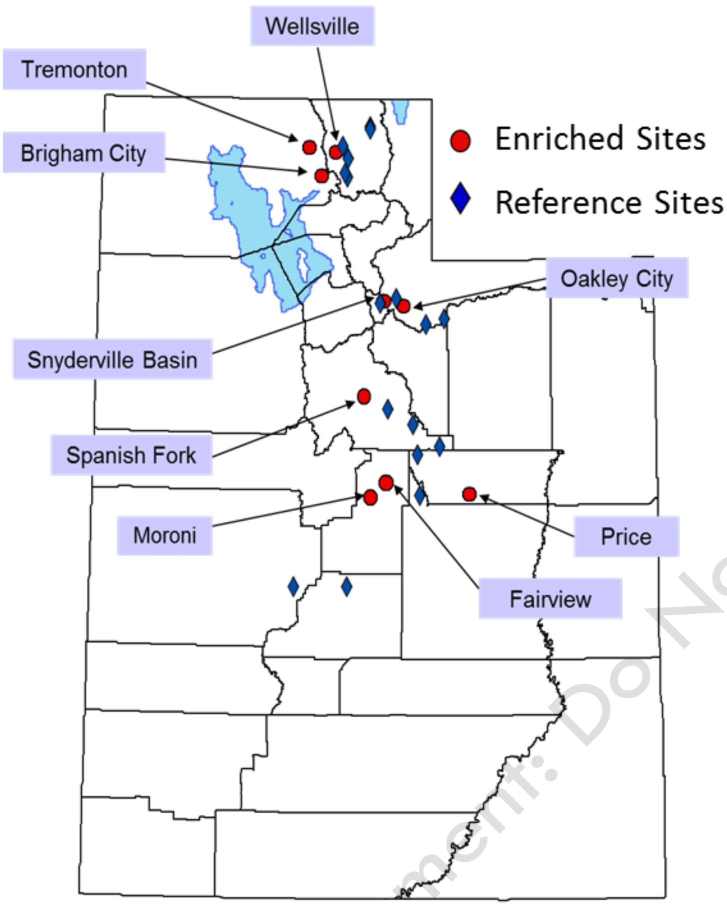


Figure 2.1. Site locations design for the 2010 ecological impacts of nutrients study. Enriched sites are a combination of two study reaches above and below a POTW discharge (labeled). Reference sites represent minimally enriched stream locations.

Study Design: Overview and Rationale

Site Selection

Functional responses were evaluated from 35 wadeable stream sites from central to northern Utah. Our principal objective in selecting sites was to collect stressor-response data from streams that together encompass the variation of N and P observed among all Utah streams. To meet this objective, we selected 17 reference sites, along with 9 sites immediately above and 9 sites immediately below the mixing zone of Publically Owned Treatment Works (POTW) effluent discharges (Figure 2.1). We attempted to measure all responses at all sites. However, in some cases field logistics precluded measurement of some indicators, so the number of sites evaluated differs slightly for different indicators (see Chapters for details).

Our site selection rationale was that upstream locations would generally be representative of streams predominantly enriched by non-point source nutrient inputs, whereas downstream locations would also include point source inputs. The intent of our sampling design was not to target POTW sources. UDWQ recognizes that these reclamation facilities are not the only, nor always the most important, sources of stream nutrients. Whenever possible, we used existing data to select reference sites with similar geomorphic characteristics (i.e., slope, width, substrate size) as the sites upstream and downstream of POTWs (Table 2.1).

Table 2.1 Streams that were sampled in the functional assessment study. Waterbody and location describe the catchment and specific locations within the catchment respectively. Code and STORET are abbreviated site designations and are included here for later reference. The Type column distinguishes reference sites from those located upstream (Above) and downstream (Below) of the POTWs.

Waterbody	Location	Code	STORET	Type
Blacksmith Fk	at U101 Xing	BLACKFK	4905440	Reference
Box Elder Ck	AB Brigham City WWTP	BEC-AB	4901180 B	Above
Box Elder Ck	BL Brigham City WWTP	BEC-BL	4901180 D	Below
Diamond Fork	BL Palmyra Campground	DIAFK	4995665	Reference
Dry Creek	AB Spanish Fork WWTP	DCSP-AB	4996020 B	Above
Dry Creek	BL Spanish Fork WWTP	DCSP-BL	4996020 D	Below
Fish Ck	AB Scofield	FISHCK	5931650	Reference
Huntington Ck	0.5 Mi BL Guard Station	HUNTCK	4931230	Reference
Little Bear R	W of Avon at Rd Xing	LBRAVON	4905700	Reference
Little Bear River	AB Wellsville Lagoons	LBRW-AB	4905600 B	Above
Little Bear River	BL Wellsville Lagoons	LBRW-BL	4905600 D	Below
Logan R	at 1000 West	LOGR1000	4905140	Reference
Logan R	By the Dugway	LOGRDUG	4905260	Reference
Logan River	BL Twin Bridges	LOGRTB	4905195	Reference
Malad R	AB Tremonton WWTP	MRTRE-AB	4902710 D	Above
Malad R	BL Tremonton WWTP	MRTRE-BL	4902710 D	Below
McLeod Ck	at Swaner Nature Preserve	KIMBALL	4925442	Reference
N Fk Chalk Ck	AB S Fk	NFCHLK	4940201	Reference
Price R	AB Price WWTP	PRP-AB	4932370 B	Above
Price R	BL Price WWTP	PRP-BL	4932370 D	Below
Price R	BL Kyune A RR Tunnel	PRICER	4932815	Reference
S Fk L Bear R	AB E Fk	SFKLBR	4905740	Reference
Salt Ck	BL Salt Canyon	SALTCK	4995355	Reference
San Pitch R	AB Fairview City WWTP	SPRFV-AB	4946830 B	Above
San Pitch R	BL Fairview City WWTP	SPRFV-BL	4946830 D	Below
San Pitch R	AB Moroni Feed WWTP	SPRM-AB	4946970 B	Above
San Pitch R	BL Moroni Feed WWTP	SPRM-BL	4946970 D	Below
Silver Ck	AB Synderville-Silver Ck WWTP	SCSNYD-AB	4926790 B	Above
Silver Ck	BL Synderville-Silver Ck WWTP	SCSNYD-BL	4926790 D	Below
Tie Fork	2 miles AB Hwy 6	TIEFK	4995928	Reference
Unknown Stream	AB Provo Falls	UKMURD	4999050	Reference
Upper Provo R	at North Fork	UPRNFK	4998700	Reference
Weber R	AB Oakley City WWTP	WROAK-AB	4928010 B	Above
Weber R	BL Oakley City WWTP	WROAK-BL	4928010 D	Below
Weber River	AB Rockport	WEBR	4927250	Reference

We assumed stressor-response relationship to be continuous among all study sites. Upstream-downstream comparisons were not conducted for several reasons. Most importantly, our primary objective was to derive general regional thresholds, as opposed to site-specific relationships. In addition, in some cases it was not clear whether direct upstream-downstream comparisons were appropriate due to locally important covariates. For instance, stream channel characteristics (i.e., width, slope) sometimes differed between upstream and downstream locations. In all cases, but to varying degrees, hydrologic characteristics differed due to the influence of the discharge. Ultimately, site-specific upstream-downstream comparisons will help inform site-specific investigations, but additional evaluations will be required to determine the extent to which physical and hydrologic factors alter observed functional responses.

Water Chemistry

We established individual study sites to be representative of stream reaches (Figure 2.2). At each reach UDWQ monitoring staff collected water chemistry grab samples at the top and bottom of the stream reaches. Depending on the potential for other nutrient inputs (i.e., irrigations return ditches, inflows), additional collection stations were established downstream of each POTW largely for the purpose of constructing water quality models (Chapter 13). Water chemistry samples were collected whenever field crews were on site, which occurs on at least three times during the summer of 2010. Field crews followed UDWQ SOPs to collect separate samples for total and filtered nutrient analyses. Samples for dissolved analytes were field filtered through 0.45 µm membrane filters (Millipore Corporation). We preserved samples by freezing them immediately following collection. The Utah State Biogeochemical Laboratory thawed samples and immediately conducted nutrient chemical analyses for total nitrogen, total dissolved nitrogen, total phosphorus and total dissolved phosphorus with persulfate oxidation followed by standard colorimetric analysis (Valderrama 1981). The lab also processed these samples with standard colorimetric analysis to obtain nitrate+nitrite (USEPA method 353.4), ammonium (USEPA method 349), and phosphate (USEPA method

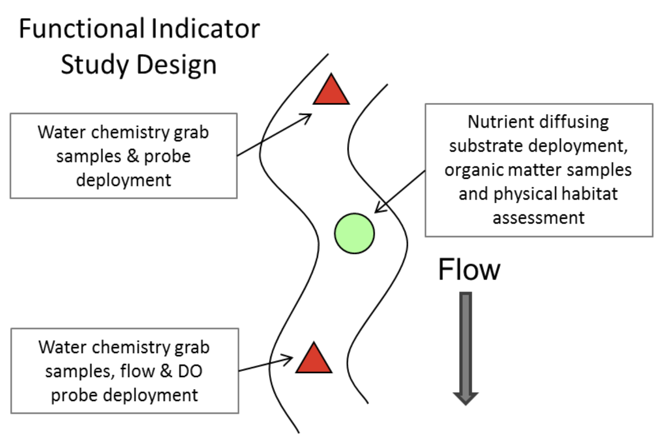


Figure 2.2. Study design for the 2010 ecological impacts of nutrients study to develop functional indicators of nutrient pollution. Sampling block design was repeated at all site locations.

365.5). Unless otherwise noted, nutrient data throughout the report are the summertime average of a minimum of 6 spatial and temporal composites for each reach.

Other water chemistry constituents, including dissolved metals and major anion/cations, were also obtained from the Utah State Health Laboratory, some of these data were used to evaluate the influence of other factors on stressor-response relationships, but most of these data will be used in follow-up site specific investigations to evaluate the potential influence of other potential stressors.

For the purpose of this report, covariate refers to any secondary variable that affects—either positively or negatively— the relationship between nutrients and ecological responses. Sometimes covariates are result from natural environmental gradients (e.g., channel shading, stream gradient). In other circumstances, covariates may arise from stressors, other than nutrients, that cause similar ecological responses to occur. The former can be accounted for in S-R relationships, whereas the latter complicates cause-effect conclusions (see Chapter 11).

Physical and Hydrologic Characteristics

Each site was surveyed using Utah's Comprehensive Assessment of Stream Ecosystem (UCASE) protocols (USEPA 2007), which includes quantitative and qualitative measures of riparian and in-stream physical and hydrological characteristics. We followed standard UDWQ procedures to obtain discharge measurements concurrent with water chemistry samples. We then calculated discharge using the velocity-area method. Velocity was measured with FlowTracker Handheld Acoustic Doppler Velocity (ADV) meters (SonTek®). Physical habitat data consisted of numerous measures of channel morphology, including wetted and bankfull widths, gradient, substrate size, and bank angles. We also measured channel shading at the banks and center of the channel at 11 equidistant transects along a stream length of 40X the wetted width of each stream.

The hydrologic and physical data were primarily used to assist with *post hoc* evaluation of the extent to which the streams in this investigation were generally representative of Utah streams and also to evaluate the potential influence of covariates on stream stressor-response relationships. Importantly, covariate analyses were broad and only accounted for their general influence on indicators among all study sites. In practice, the influence of these covariates is intrinsically site-specific. In other words, these analyses can tell us if something like slope generally influences a response metric, but it cannot tell us the extent of this influence for a specific site.

Deployment of Multi-Parameter Sondes

Field crews deployed water quality probes (YSI 6600V2 or 600 OMS V2) for a minimum of 48-hours at each upstream and downstream station during the peak growing season (July through early September). All sondes recorded measures of pH, temperature, dissolved oxygen (DO), and specific conductance every 5 minutes. The larger sondes (YSI 6600V2) also recorded chlorophyll and turbidity concentrations. All parameters were calibrated simultaneously on site and immediately prior to deployment. Independent measures, from a recently calibrated sonde, were collected at the end of each deployment to evaluate whether sensor fowling resulted in parameter drift.

To improve the accuracy of calculations, two-station stream metabolism sondes were placed at the top and bottom of each reach with optimal between sonde spacing—the approximate distance required for half of the DO molecules to exchange with the atmosphere. Estimates of optimal sonde spacing were derived from the following relationship (Grace and Imberger 2006):

$$D = \frac{v^{0.33}}{50.8 \times (d^{-0.85})}$$

where:

D = probe distance (km)
v = stream velocity (cm/sec)
d = mean water depth (cm)

Nutrient Diffusing Substrates

We also deployed nutrient diffusing substrates (NDSs) in a glide located approximately midway within each reach. NDSs were positioned within each glide at a location that maximized exposure to sunlight (typically the center of the channel).

Organic Matter Standing Stocks

We collected volumetric samples of various stores of organic matter standing stock along a minimum 50-meter stream segment approximately midway within each reach. We collected these samples at a separate event at the end of the field season.

Discussion

Mean nutrient concentrations among study sites varied considerably for both TN (0.11-14.72 mg/L) and TP (0.003- 7.89 mg/L) (Table 2.1). For context, we compared these TN and TP concentrations these sites with statewide estimates and the nutrient concentrations among our study sites span the range of predicted statewide observations (Figure 2.3). We purposefully sampled a disproportionate number of sites with unusually high and low (reference site) nutrient concentrations because these sites are relatively rare and we wanted to avoid a single site having too much leverage in regional stressor-response relationships. We assumed that selecting representative populations of streams with relatively low, moderate and high nutrient concentrations provides better estimates of the range of functional responses that we might encounter among Utah streams. However, these streams also differed with respect to other characteristics with the potential to confound stressor-response relationships. As a result, our analyses also evaluated the effect of covariates on functional responses. It is important to note that the inclusion of streams with diverse characteristics, located

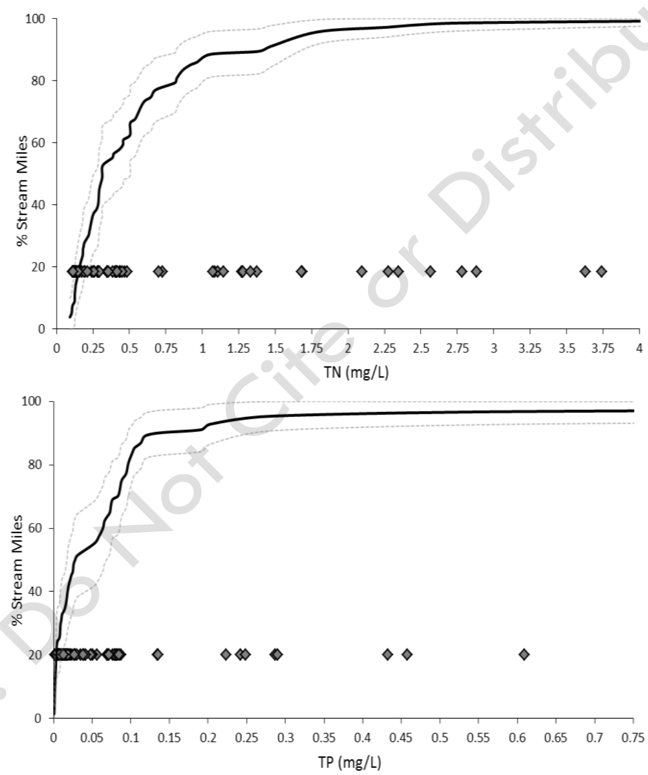


Figure 2.3. Solid black lines represent the cumulative frequency distribution from randomly selected sites throughout Utah, whereas dashed lines depict the 95th percent confidence interval of these relationships. Grey diamonds are the average nutrient concentrations obtained from the sites used in this functional response study. These plots do not include three high TN and four high TP sites because they exceeded the plot scale.

throughout the state, helps guard against over fit stressor-response models. Missing important covariates would generally weaken, as opposed to strengthen, regional stressor-response models due to systematic bias (Yuan et al. 2010). In other words, unaccounted for covariates would generally obscure linkages between nutrients and responses. On the other hand, depending on the specific response under evaluation, such systematic bias does obscure the ability for stressor-response models to demonstrate that nutrients are the principal cause of any deleterious responses observed. UDWQ will continue to conduct more detailed covariate analyses For instance, we already more thoroughly evaluated the influence of covariates on headwater N and P concentrations in anticipation of NNC development (Chapter 8) and we have outlined tools that could be applied to NNC development on a site-specific basis.

Table 2.2 Nutrient concentrations and physical characteristics of all upstream and downstream sites and representative reference sites that were sampled for the functional response study.

Watershed	Location	TN (mg/L) Mean	TN (mg/L) SD	TP (mg/L) Mean	TP (mg/L) SD	Area (mi^2)	Forested (%)	Slope (%)	Channel Shading (%)	Mean Depth (cm)	Mean Width (m)	Discharge (cfs)
Box Elder Creek	above Brigham City POTW	0.404	0.117	0.058	0.038	37.9	10.5	0.3	25.0	11.1	2.99	0.3
Box Elder Creek	below Brigham City POTW	6.351	4.346	0.838	0.716	38.1	10.5	0.1	92.1	30.5	5.17	2.7
Blacksmiths Fork	at HWY 101	0.188	0.040	0.008	0.001	269.0	31.1	1.3	77.3	40.2	8.26	116.1
Dry Creek	above Spanish Fork POTW	2.216	0.895	0.135	0.063	18.1	35.9	0.3	14.8	45.6	4.43	2.7
Dry Creek	below Spanish Fork POTW	11.286	3.083	1.894	0.470	19.2	34.2	0.2	0.0	72.0	5.75	4.2
Diamond Fork	at Palmyra Campground	0.410	0.063	0.084	0.041	139.0	57.4	1.7	5.6	24.2	11.69	68.4
Fish Creek	wbove Scofield Resevoir	0.246	0.049	0.063	0.072	63.2	67.4	0.3	11.0	14.9	9.98	5.9
Huntington Creek	in Huntington Canyon	0.455	0.063	0.015	0.003	46.0	69.8	1.5	16.6	20.9	6.09	18.3
Kimball Creek	at Swaner Nature Preserve	0.279	0.057	0.028	0.004	29.9	68.7	0.3	8.6	25.2	4.03	15.2
Little Bear River	west of Avon	0.343	0.234	0.021	0.009	77.6	21.7	0.8	34.1	16.7	7.86	3.5
Little Bear River	above Wellsville Lagoons	1.175	0.183	0.075	0.014	250.0	16.7	0.2	96.6	21.7	7.85	9.6
Little Bear River	below Wellsville Lagoons	1.085	0.163	0.084	0.010	255.0	16.5	0.2	5.2	60.4	8.62	27.9
Logan River	at 1000 W	0.424	0.080	0.023	0.005	255.0	33.6	0.3	69.7	41.8	11.27	48.0
Logan River	below the Dugway	0.139	0.030	0.012	0.002	135.0	40.3	1.9	16.0	27.4	11.99	116.1
Logan River	below Twin Bridges	0.132	0.025	0.012	0.004	141.0	40.1	1.3	19.7	29.3	10.63	73.0
Malad River	above Tremonton POTW	2.828	0.570	0.236	0.051	689.0	5.2	0.1	41.1	57.4	8.45	26.0
Malad River	below Tremonton POTW	3.897	1.204	0.445	0.144	691.0	5.2	0.1	37.0	65.8	9.07	27.0
N Fk of Chalk Creek	above South Fork confluence	0.161	0.040	0.006	0.003	27.5	81.6	3.4	80.0	11.3	5.1	13.5
Price River	below Kyune	0.388	0.048	0.050	0.014	327.0	61.6	0.9	4.0	33.6	12.65	103.1
Price River	above Price POTW	0.710	0.136	0.289	0.368	720.0	44.6	0.3	11.6	34.5	6.26	21.3
Price River	below Price POTW	2.950	1.097	0.732	0.293	849.0	42.1	0.2	12.3	32.5	7.27	19.0
Salt Creek	above mouth of Salt Canyon	0.154	0.112	0.016	0.004	95.1	48.3	2.5	71.0	20.6	2.55	3.9
Silver Creek	above Snyderville Basin POTW	0.319	0.121	0.015	0.008	15.9	40.2	0.2	10.7	47.2	1.60	1.1
Silver Creek	below Snyderville Basin POTW	14.717	4.613	2.212	0.715	17.2	38.9	0.5	1.5	45.0	2.41	2.0
S Fk of Little Bear River	below Davenport Creek	0.229	0.040	0.017	0.003	63.0	20.9	1.6	25.4	23.3	8.19	15.8
Spanish Fork River	above Fairview POTW	1.324	0.215	0.019	0.009	96.5	56.5	0.6	4.4	20.2	5.08	5.5
Spanish Fork River	below Fairview POTW	1.677	0.232	0.078	0.035	98.0	55.9	0.5	10.3	32.1	4.25	6.7
Spanish Fork River	above Moroni POTW	1.232	0.116	0.033	0.008	227.0	42.9	0.3	6.9	31.4	5.77	0.8
Spanish Fork River	below Moroni POTW	10.416	1.194	7.897	1.449	228.0	42.6	0.2	20.7	41.6	6.40	2.0
Tie Fork	two miles above HWY 6	0.118	0.011	0.007	0.002	16.8	64.6	2.4	24.6	15.4	2.30	2.5
Unnamed Stream	in Murdock Basin	0.109	0.017	0.004	0.001	3.7	66.5	4.7	48.9	14.6	3.00	0.8
Provo River	at North Fork Trailhead	0.113	0.023	0.003	0.003	66.6	77.8	7.1	4.3	24.5	13.94	54.4
Weber River	above Rockport Resevoir	0.382	0.068	0.040	0.012	175.0	71.8	1.3	12.6	49.3	17.04	116.1
Weber River	above Oakley City POTW	0.109	0.023	0.009	0.002	173.0	72.1	1.0	12.6	28.7	12.69	35.1
Weber River	below Oakley City POTW	0.125	0.040	0.020	0.017	175.0	71.6	1.1	20.1	33.7	11.19	35.8

CHAPTER 3

EXPERIMENTAL ESTIMATES OF NUTRIENT LIMITATION AND SATURATION

Introduction

Numerous studies indicate that the growth and productivity of autotrophic biota in aquatic systems are often limited by nutrient availability, most commonly nitrogen (N) and/or phosphorus (P) (see Francouer 2001 and Elser et al. 2007 for reviews). Within a stream, nutrient availability is related to autotrophic production (Fairchild et al. 1985, Bernhardt and Likens 2004), as well as whole ecosystem primary production (Mullholland et al. 2001, Elser et al. 2007, Bernot et al. 2010). This relationship has been extensively described in lakes, where the supply of N and especially P are strongly related to phytoplankton biomass (Schindler 1977, Hecky and Kilham 1988, Elser et al. 1990). Others have also demonstrated the importance of N and P in predicting stream algal biomass (Welch et al. 1988, Chatelot et al. 1999, Dodds et al. ,1997, Biggs 2000, King et al. 2000), although relationships are less generalizable

Across broader scales, however, the relative effects of different water column nutrient concentrations (i.e., N:P ratios) on benthic algal growth is more complex. For instance, when comparing N:P ratios to chlorophyll *a* measurements across multiple stream systems, Francoeur (2001) and Johnson et al. (2009) observed both positive and negative correlations between water column nutrients, physical stream features, and benthic algae growth (measured as chlorophyll *a*), i. Important physical factors with the potential to limit aquatic primary production include light availability (Bernhardt and Likens 2004 and Hill and Fanta 2008), scouring or substrate stability (Mulholland et al. 1991), and temperature (Sanderson et al. 2009). In larger soft-bedded rivers the sorption and subsequent release of nutrients from stream sediments is another important factor. In rivers with extensive macrophyte beds, nutrient uptake can be rapid and retention longer because few organisms directly feed on plant tissue (Riis et al. 2012). Other biotic interactions, such as competition with heterotrophs (Tank and Dodds 2003) and invertebrate grazing (Rosemond et al. 2000, Opsahl et al. 2003), also alter rates of primary production in stream ecosystems.

The relative importance of different controls on stream algal production is of keen interest to water resource managers. Knowing which nutrients, if any, limit algal growth helps inform resource managers about specific mitigation actions that are most likely to improve or maintain the biological integrity of aquatic ecosystems. This is particularly true for point sources with treatment facilities that have the capability of removing either N or P. Whether or not a facility treats for N, P, or both can have appreciable ramifications to operating costs (e.g., see CH2MHill 2009). Nonpoint sources are more difficult to track and control, but nevertheless knowing which nutrient is limiting helps to inform the selection of specific best management practices (BMPs) that are most likely to result in water quality improvements. These different nutrient sources are also potentially important in the alteration of limitation patterns, because NPS nutrient sources are episodic, whereas point sources are much more constant.

One way to measure nutrient limitation is the use of bioassays. In streams, the most common approach is the use of nutrient diffusing substrates (NDS,) (Tank and Dodds 2003, Johnson et al. 2009, Scott et al. 2009), in part because of the relatively low cost of these experiments. NDSs are composed of an agar medium infused with a single nutrient or combination of nutrients (most commonly N, P, and/or both) that diffuse through a solid porous substrate into the water column. The solid porous substrate serves as a platform for algal colonization, and can also be easily removed to measure algal biomass accrual. With these experiments, it is assumed that the nutrient (or combination of nutrients) treatment that leads to the greatest increase in algal biomass—typically expressed as Ash Free Dry Mass (AFDM) or chlorophyll-a per area of stream bed—is mostly likely to limit benthic production.

In this study, Utah's Division of Water Quality (UDWQ) used NDS to evaluate nutrient limitation to benthic primary production at 35 streams throughout central and northern Utah. The objectives of this study were to: 1) determine the limiting nutrient for algal growth in least-disturbed reference sites across Utah, 2) evaluate the effects of moderate and high levels of eutrophication on nutrient limitation, and 3) quantify the TN and TP concentrations where saturation with respect to NDS biomass accrual generally occurs. Together these measures provide insight into the relative importance of N and P in benthic primary production.

Methods

Study Sites

During the summer of 2010 NDS bioassays were deployed in 29 of the 35 sites that were included in the functional response study. Of these sites, 15 were in reference condition and relatively uninfluenced by human-caused perturbations. The remaining 14 study sites were located upstream (n = 7) and downstream (n = 7) of Publically Owned Treatment Works (POTWs) to encompass the range of ambient stream nutrient among streams throughout Utah and a variety of nutrient sources (see Figure 2.3, Chapter 2). Deployments took place from July to August during baseflow conditions and when potential algal growth rates were greatest due to higher temperatures and longer days.

For the purpose of NDS analyses, we created a priori site classes based on differences in nutrient sources. A summary of these *a priori* site categorizations follows:

Reference Sites: Minimal effects of human activities, including nutrient inputs.

Moderately Enriched Sites: Primarily nonpoint source nutrient inputs, variable but sometimes extensive habitat modifications.

Highly Enriched Sites: Both point source and nonpoint source nutrient inputs, variable but sometimes extensive habitat modifications.

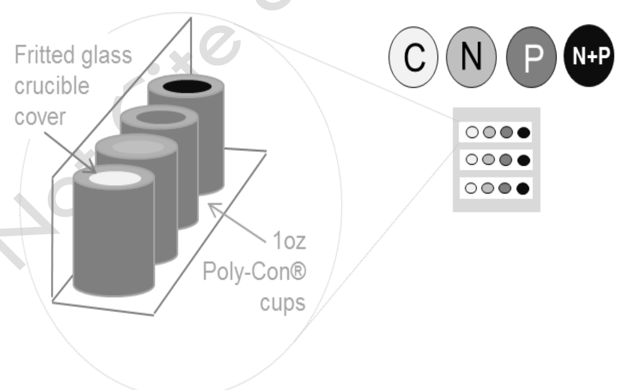


Figure 3.1. Diagram of NDS housing and the configuration of treatments.

Nutrient Diffusing Substrates

To assess nutrient limitation on algal growth, UDWQ constructed NDS housing and nutrient enriched agar, using procedures developed by Utah State University's (USU) Aquatic Biogeochemistry Laboratory, which followed the methods described in Tank et al.'s (2006) procedures in *Methods in Stream Ecology*. NDS units were constructed from 1oz Poly-Con® cups (Median Plastics) with a ¾ inch hole drilled into a hinged lid. We then filled the cups filled with approximately 30 mL of agar solution with different combinations of nutrients to obtain four different nutrient treatments: control (Agar), nitrogen (N as NH_4Cl, 0.5 mol/L), phosphorus (P as KH_2PO_4, 0.5 mol/L), and nitrogen + phosphorus. Finally, we topped each NDS housing unit with a 2.75 cm diameter fritted glass crucible cover (Leco, Inc) to provide an inorganic surface for periphyton to grow (Figure 3.1). To facilitate stream deployment, we constructed a housing unit for the nutrients treatments from a 12 inch plastic L-shaped bar (US Plastic, part 48445), and then secured to a cinder block. Each cinder block held 3 replicates for each of the four nutrient treatments. Cinder blocks were deployed perpendicular to streamflow in a representative glide within each stream.

We developed deployment protocols to minimize variation in other factors (i.e., shading and scour) that could potentially obscure among stream differences in NDS patterns. At each site, we selected a deployment location in a representative flowing water habitat (i.e. glide or run). NDS deployed upstream of POTW discharges were positioned above, but as close as possible, to the POTW discharge location. Below POTWs NDS were placed directly below the mixing zone of the effluent and receiving waters. To determine the mixing zone, we measured specific conductance at ≥5 points across the stream channel (YSI data sondes) at several locations downstream. We continued these measurements until conductivity measurements across the width of the stream were homogenous, and selected the next downstream glide or run as the deployment site.

Once representative glides or runs were established, we selected a specific deployment location in an area that maximized sun exposure, which helped minimize, to the greatest extent possible, among stream differences in channel shading. We then deployed NDS cinderblocks perpendicular to the flow to minimize among treatment contamination. Following standard deployment protocols, we incubated NDS bioassays for ~21 days (Tank et al. 2006). We inspected NDS housings twice during the incubation period, and cleared any visually observable settled organic material or entangled debris (e.g., grasses or small sediments debris) that had accumulated on the cinder block and/or NDS housing units. After the incubation period, we removed the crucible covers and immediately froze them to preserve the samples until immediately prior to laboratory chlorophyll analyses.

Water Chemistry

We collected surface water grab samples of 50-ml upstream and immediately downstream of the NDS three times during the NDS incubation period. Samples were kept frozen until analyzed by USU's Aquatic Biogeochemistry Laboratory for TN and TP (see Chapter 2 for details, Valderrama 1981). Chlorophyll *a* analyses were conducted by the State of Utah Public Health Laboratory with standard flourometric methods—uncorrected for pheophytin (USEPA Method 445.0).

Data Analysis

QUANTIFICATION OF EXPERIMENTAL RESPONSES

We quantified among-treatment differences in algal accrual by measuring the relative differences in chlorophyll *a* (*chl-a*) that accumulated on each fritted glass cover during deployments. In most cases, NDS *chl-a* results (mg/m^2) are expressed as an absolute 21-day accrual. However, in several instances (n=8) deployments lasted slightly longer than 21 days due to field logistics, which necessitated use of a linear correction to standardize among stream comparisons.

We used two-factor analysis of variance (ANOVA) to test whether *chl-a* increased on N and/or P treatments. Prior to ANOVA analyses, we square root (x) transformed *chl-a* concentrations to meet statistical normality assumptions. We followed generally held statistical conventions and set statistical significance (alpha) at $p<0.05$. At 11 sites a replicate sample was lost, in which case we equally weighted the remaining replicates in order to balance the ANOVA. We then translated the statistical outcomes of these tests into ecologically meaningful nomenclature following the work of Tank and Dodds (2003), where:

- ***Single nutrient limitation*** occurs when N or P (but not both) elicit an increase in NDS algae accrural.
- N and P are ***independently colimited*** when neither N nor P treatments elicits an increase in bethic algae, but the addition of both N and P increases NDS algae accrual.
- N and P can also be ***colimited*** when N and P additions both elicit a positive response, even if the addition of both does not result in additional algae accrual.
- It is also possible that NDS algae accrual is ***not limited by either N or P***—for instance if production is limited by another factor (i.e., light, micronutrient). In this case, neither N nor P nor the two nutrients together elicit an increase in NDS primary production.

SATURATION THRESHOLDS

We were also interested in determining whether we could identify in-stream concentrations of TP and TN where nutrients were saturated (i.e., no longer limiting to algal growth). For these analyses we first divided sites into two groups: those that showed nutrient limitation and those sites that did not. Next, we used non-parametric deviance reduction (NDR) (Qian et al 2003, package `rpart`) with water column nutrients as the independent variable and nutrient limitation as a binary, dependent variable (e.g., any or no nutrient limitation) to establish N and P thresholds that best delineate these groups. To test the significance of the thresholds we used a linear mixed model (package `lme4`) for sites on each side of the threshold to determine if there were any differences in limitation. Pairwise differences among groups was subsequently evaluated with ANOVA followed by Tukey's Honestly Significant Difference (HSD) post hoc tests.

One potential problem with NDR approaches is that derived thresholds are sometimes overfit because they are overly dependent on the specific dataset used to generate them. To determine if our results were generalizable beyond the dataset used to derive thresholds, we used Receiver

Table 3.1. Nutrient limitation determined by NDS deployments and ambient water column nutrient concentrations for each site in the study . Treatment effect was determined by a two-way ANOVA. Site descriptions can be found in Chapter 2, Table 2.1.

	Nutrient	TN (mg/L)		TP (mg/L)		
Site Code	Limitation	Mean	SD	Mean	SD	Site Type
BLACKFK	N+P	0.188	0.040	0.008	0.001	Reference
DCSP_AB	None	2.216	0.895	0.135	0.063	Moderately Enriched
DCSP_BL	None	11.286	3.083	1.894	0.470	Highly Enriched
DIAFK	N	0.410	0.063	0.084	0.041	Reference
FISHCK	N+P	0.246	0.049	0.063	0.072	Reference
KIMBALL	N+P	0.279	0.057	0.028	0.004	Reference
LBRAVON	None	0.343	0.234	0.021	0.009	Reference
LBRW_AB	N	1.175	0.183	0.075	0.014	Moderately Enriched
LBRW_BL	None	1.085	0.163	0.084	0.010	Highly Enriched
LOGR1000	None	0.139	0.080	0.012	0.005	Reference
LOGRDUG	None	0.424	0.030	0.023	0.002	Reference
LOGRTB	N+P	0.132	0.025	0.012	0.004	Reference
MRTRE_AB	P	2.828	0.570	0.236	0.051	Moderately Enriched
MRTRE_BL	N+P	3.897	1.204	0.445	0.144	Highly Enriched
NFCHLK	N+P	0.161	0.040	0.006	0.003	Reference
PRICER	N+P	0.388	0.048	0.050	0.014	Reference
SCSNYD_AB	N+P	0.319	0.121	0.015	0.008	Moderately Enriched
SCSNYD_BL	None	14.717	4.613	2.212	0.715	Highly Enriched
SFKLBR	N	0.229	0.040	0.017	0.003	Reference
SPRFV_AB	None	1.324	0.215	0.019	0.009	Moderately Enriched
SPRFV_BL	None	1.677	0.232	0.078	0.035	Highly Enriched
SPRM_AB	P	1.232	0.116	0.033	0.008	Moderately Enriched
SPRM_BL	None	10.416	1.194	7.897	1.449	Highly Enriched
TIEFK	N+P	0.118	0.011	0.007	0.002	Reference
UKMURD	N	0.109	0.017	0.004	0.001	Reference
UPRNFK	N+P	0.113	0.023	0.003	0.003	Reference
WEBR	N	0.382	0.068	0.040	0.012	Reference
WROAK_AB	N+P	0.109	0.023	0.009	0.002	Moderately Enriched
WROAK_BL	N+P	0.125	0.040	0.020	0.017	Highly Enriched

Operator Characteristics (ROC, package `pROC`). ROC evaluates the performance of a binary classification system using bootstrapped data to generate several metrics of model performance. We evaluated the strength of thresholds identified by NDR with several of these metrics, including: the Area Under the Curve (AUC), sensitivity (true positive rate), and the specificity (true negative rate) for the TN and TP thresholds. AUCs quantify the accuracy of models by calculating the probability that at the threshold established a randomly chosen site has the predicted response given previously defined "acceptable" Type I and II error rates. In this case AUCs provide the percent of the measurements for which our thresholds correctly identified sites saturated by nutrients. Other measures of the performance of these threshold models included sensitivity and specificity with the following equations:

Sensitivity = True Positives/True Positives + False Negatives
Specificity = True Negatives/True Negatives + False Positives

All analyses were conducted in R v2.15.0 (R Core Development Team, 2012).

Results

Enrichment Classes

We observed significant differences in TN and TP among the three nutrient enrichment classes (ANOVA, $p<0.001$):

- Reference Sites: n=15, TN 0.25 and TP 0.027 mg/L
- Moderately Enriched Sites: n=7, TN 1.16 and TP 0.098 mg/L)
- Highly Enriched Sites: n=7, TN 6.57 and TP 1.72 mg/L)

These differences allowed us to subsequently evaluate the extent to which NDS nutrient limitation varied systematically with nutrient enrichment.

N vs P Limitation among Utah Streams

Among all sites, we observed several patterns of nutrient limitation: no limitation, sole nitrogen or phosphorus limitation and colimitation (Table 3.1). The most common condition observed among all sites was colimitation of N and P, which we observed at 12 sites. The next most common pattern was nutrient saturation—no increase in chlorophyll concentrations in response to nutrient additions—which we observed at 10 sites, half of which were within the highest enrichment class. Relatively few sites (n=7) were limited by a single nutrient, and among these sites N-limitation (n=5) was more common than P-limitation (n=2).

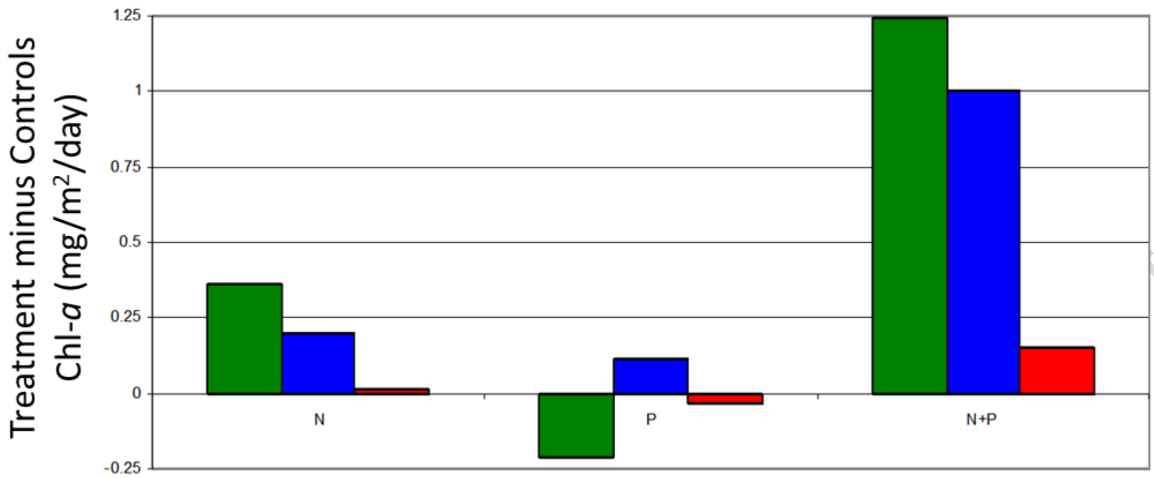


Figure 3.2 Accrual of chlorophyll for treatments relative to controls among stream within each of the three nutrient classes: reference(green), moderately enriched (blue) and highly enriched (red) streams.

General Limitation Patterns: Nutrient Treatments vs. Controls

With the exception of reference site phosphorus treatment, NDS algae accrual was, on average, predictably higher for nutrient treatments relative to controls. The magnitude of treatment responses differed among enrichment classes, with reference sites showing the greatest response followed by moderate then highly enriched streams (Figure 3.2). N generally had greater control on NDS algae accrual than P. This was particularly for true treatments that augmented both N and P. On average, the addition of both N and P resulted in as much as 3X more algae accrual than the addition of P alone. While found this pattern among all nutrient enrichment classes, but the magnitude of the response generally decreased with increasing ambient nutrient concentrations. In fact, many of the streams with the highest background nutrient concentrations did not respond to any of the nutrient treatments.

Table 3.2. Mean nutrient concentrations (±standard deviation) and the significance of the treatments within each enrichment class. Statistical significance of treatments was determined with linear mixed models where X indicates significance at $p<0.05$ and x* indicates significance at the $p<0.1$.

Enrichment Class	Sample Size	Mean Nutrient Concentration (mg/L)		None	N	P	NxP
		TN	TP				
Reference	15	0.25±0.12	0.027±0.025		X		X
Moderately Enriched	7	1.16±0.93	0.098±0.010	X	X*		
Highly Enriched	7	6.57±5.66	1.72±2.560	X			

Limitation Patterns among Enrichment Classes

Patterns of limitation also differed among nutrient classes. We observed significant chlorophyll increases in treatments relative to controls among reference sites for both N (Mixed Linear Model, $p=0.001$) and N and P treatments ($p<0.001$, Table 3.2). However, enriched sites were not consistently N, P or colimited. In fact, when sites were binned by limitation, the only statistically distinct pattern among streams in both enrichment classes was a general tendency toward no limitation by macronutrients. In total 71% of streams within the most highly enriched class did not respond to any nutrient treatment.

NUTRIENT LIMITATION AT REFERENCE SITES

Among all reference streams, we found that 80% to be limited by one or more nutrients. The most common limiting condition found at reference sites was colimitation, which occurred at 53% (8/15) of reference streams. Patterns among the remaining reference sites included N-limitation which occurred at ~26% (4/15 sites). Surprisingly, three reference sites (20%) were not limited by macronutrients and no reference site was solely P limited (Table 3.1).

NUTRIENT LIMITATION AT ENRICHED SITES

Limitation patterns were variable among moderately streams within enrichment classes. While most highly enriched sites did not respond to any nutrient treatment, among the streams that did one was N-limited and two were co-limited. Sites with moderate levels of nutrient enrichment were most variable and exhibited every type of nutrient limitation is roughly equal proportion: 26% (2/7 sites) were limited by P, 26% by N and P, 26% did not respond to treatment, while 14% (1/7) were limited by N.

Saturation Thresholds

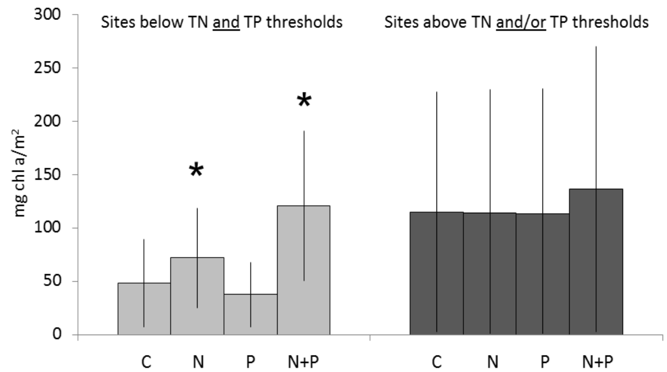


Figure 3.3. Chlorophyll *a* concentrations by treatment for sites below the TN and TP thresholds (left panel, 0.42 and 0.078 mg/L, respectively, n=16) or above either threshold (right panel, n=13). Error bars are one standard deviation and asterisk indicate a significant treatment effect using a linear mixed model.

We used deviance reduction analysis to identify ambient nutrient concentrations associated with sites where the N and P treatments did not increase algae growth (i.e., saturation thresholds) as opposed to those that exhibited an experimental response. For N, we found that above a threshold of 0.42 mg/L TN sites were more likely to be saturated (95% CI, 0.33 -1.4 mg/L), whereas TP saturation thresholds occurred a concentration of 0.078 (95% CI, 0.017 -1.33 mg/L). In both cases, we further confirmed that the thresholds were reasonably robust with Receiver Operating Characteristic (ROC) models. Both models were reasonable accurate with an AUC of 81.7% (95% CI, 64.0-99.4%) for TN thresholds and 71.9% (95% CI, 51.0-92.9%) for TP, which indicates that sites would generally be correctly classified into these groups about 70-80% of the time. Other measures of accuracy were equally promising with sensitivity (TN = 0.78 (95% CI, 0.44-1.0), TP = 0.56 (95% CI, 0.23-0.89)) and specificity (TN = 0.75 (95% CI, 0.55-0.90), TP = 0.80 (95% CI, 0.60-0.95)) both suggesting a balance between Type I and II errors.

We evaluated patterns of nutrient limitation among sites above and below the TN and TP saturation thresholds and found that, on average, sites below the thresholds show similar patterns as those observed among reference sites. Liner mixed models revealed that sites with ambient TN and TP below saturation thresholds had significant N limitation ($p<0.001$) and colimitation ($p<0.001$). In contrast, sites with ambient nutrient concentrations above both thresholds did not respond to any of the nutrient treatments, which indicates that these sites were at saturation points with respect to TN and TP (for all treatments $p>0.37$, Figure 3.3).

Discussion

For more than a decade the U.S. Environmental Protection Agency (USEPA) has emphasized the importance of States developing NNC to protect beneficial uses (USEPA 2010). Development of protective nutrient criteria requires an understanding of the relationship between nutrients and the deleterious effects that nutrient enrichment can cause in a waterbody. Increased periphytic algal productivity is among several import primary responses to increased nutrients, and can lead to decreased aesthetic values (Suplee et al. 2009, also see Chapter 7 in this report) and/or contribute to large dissolved oxygen daily fluctuations with low nighttime minima that can be harmful to macroinvertebrates and fish (aquatic life uses, see Chapter 6). This study, among others (i.e., Elsner 2000 metanalysis), consistently demonstrate that streams are highly variable with respect to nutrient limitation, so understanding the relative role of N and P is determining algal accrual is useful in the context of meeting nutrient reduction management objectives. This study demonstrates the potential utility of NDS bioassays as a potentially useful tool in providing resource managers with this information.

Limitations of Bioassay Experiments

Ecologists have long grappled with the relative advantages and disadvantages of ecological surveys vs. controlled experiments. Ecological surveys are most realistic, but can be difficult to interpret because many important ecological characteristics covary. Experiments seek to control some of these factors in order to study others, but ecosystem-scale experiments are expensive and difficult (if not impossible) to replicate and the results of smaller scale experiments (such as the NDS bioassays discussed in this chapter) cannot always be accurately extrapolated to the ecosystem of interest. Both of these approaches have merit, but it is critical that one understands the limits of any approach when interpreting the data.

In the case of these investigations, we were primarily interested in how nutrient limitation differed among streams with different ambient nutrient concentrations. As a result, our experiments controlled several other factors that influence benthic production. For instance, we conducted these experiments during the peak growing season, which assumes that nutrients would be similarly limited during other periods of the year. In fact, some investigations have documented temporal shifts in nutrient limitation (Biggs et al. 1999, Sanderson et al. 2009), which have been attributed to seasonal changes in temperature, temporal fluctuations in ambient nutrient concentrations or stoichiometry (Francoeur et al. 1999), or seasonal changes in light availability (Rosemond et al. 2000).

Another important influence on NDS nutrient limitation is light (Von Shiller et al 2000, Francoer et al. 1999). We attempted to control for among stream differences in channel shading by deploying the NDS experiments in areas of the stream with minimal channel shading, which assumes that patterns of nutrient limitation are similar under both low and high light conditions. This choice also means that we may have underestimated saturation concentrations, but this would only be the case if many of our sites had both high nutrients and an extensive canopy cover, which rarely occurred. This decision also means that we likely underestimate the number of streams where light, as opposed to nutrients, is the factor most limiting benthic algae growth.

Another critical assumption is that the algae species that grow on the fritted glass filters have similar nutrient requirements as algae species in other stream habitats. The evidence of the extent to which these compositional differences affect NDS results is mixed. Algae species can differ in nutrient requirements. For instance, in one study McCormick et al. (2001) observed N-limitation for the macroalgae *Cladophora* in a stream that was generally P-limited. Similarly, algae assemblages in high nutrient environments can differ in nutrient responses from assemblages found in low nutrient environments (King et al. 2000). With regard to nutrient bioassays, changes in treatment concentrations could result from differences in the chemical of physical properties (e.g., roughness) of the glass filter relative to stream substrate or from nutrients (Stelzer and Lamberti 2001). The former has been investigated in a couple of ways. Capps et al. 2011 evaluated the effects of various NDS methods and while all methods showed that the addition of both N and P has the greatest effects, more subtle differences in patterns of co-limitation were methodological.

These experiments were intentionally general because we were exploring the potential value of NDSs as a line of evidence in support of Utah's nutrient management strategy. However, the assumptions, among others, will need to be considered on a site-specific basis before concluding, for instance, that the control of N or P is not needed to address a nutrient-related impairment. Nevertheless, the relative influence of these factors could be evaluated, in part, with systematic alteration of the experimental design of NDS bioassays experiments.

Nutrient Limitation at Reference Sites

It is useful to examine patterns of nutrient limitation among reference sites because as we observed in this investigation, limitation patterns can change as streams become degraded. Sites in reference condition are frequently used to establish water quality goals or restoration expectations (Stoddard et al. 2006). Although it is important to note that reference conditions are not always indicative of the best attainable condition, particularly at streams with irreversible habitat or hydrologic alterations. Based on our NDS study it appears that benthic algae production in Utah's reference streams is most likely to be N limited or colimited by N and P, because these conditions were observed at 80% of all reference sites. None of the reference sites that we evaluated were solely P limited.

The reference site NDS results highlight the importance of including both N and P for Utah's nutrient reduction strategy. On average, algal growth on N treatments were ~36% greater than control treatments, whereas N and P treatments were ~124% greater than the control. Additionally, the combined influence of generally increased algae responses (Figure 3.2), which highlights the importance of considering both nutrients in the context of Utah's nutrient reduction strategy. Our results from linear mixed models showed both significant N effects and an N and P interaction (Table 3.2). In such circumstances, it may be more appropriate to classify stream benthic algae growth as primarily N limited and secondarily P limited than strictly N and P colimited (Tank and Dodds 2003). At these sites, increases in N would be predicted to increase benthic algal growth initially, but the N demands of algae would be met relatively quickly requiring P increases for further algae accrual.

Colimitation of N and P appears to be the predominant natural state among Utah's streams This observation concurs with several recent reviews of freshwater nutrient limitation that stressed the importance of colimitation over the long held paradigm of single source, P limitation (Harpole et al. 2011, Lewis et al. 2011). One of these studies was a meta-analysis of experimental nutrient enrichment studies and found that, in general, additions of both N and P additions in freshwater, marine, and terrestrial environments led to a greater response than either nutrient alone (Elser et al. 2007). In addition, the magnitude of responses in freshwater ecosystems may be exponential, because the response to additions of both N and P is greater than the sum of responses to individual additions of N or P (Allgeier et al. 2011).

Although 12 of our 15 reference sites showed some form of limitation, we were surprised to find that three sites did not respond to experimental nutrient increases. These sites had low nutrient concentrations (mean 0.23 and 0.016mg/L TN and TP, respectively) that were similar to the average overall population of reference streams (mean 0.24 and 0.025 mg/L TN and TP, respectively Something other than nitrogen and phosphorous likely controls benthic algae growth rates in these streams. One possibility is that algal growth at these sites was limited by micronutrients such as Fe, Mn, Zn, Co, Mo (Pringle et al 1986, Passy 2008) instead of N or P. A second possibility—and we think most likely—is that these sites were light limited. Each of these sites had heavy shading, and while we tried to minimize shading effects at reference sites by selecting deployment locations that received the most direct sunlight, we may not have been able to eliminate site-specific shading effects at these locations. Therefore, our results demonstrate the effects of nutrients in the more sensitive stream reaches where light is not the predominant limiting factor in benthic algae accrual. These results further highlight the importance of understanding the influence of important, site-specific covariates when interpreting ecological responses to nutrient enrichment.

Patterns of Limitation among Enriched Streams

Among those enriched streams that were not saturated with N or P or limited by other factors, limitation patterns were highly variable. One possibility explanation is that the variable responses are manifestation of human-caused enrichment. For instance, the streams in this study may differ in the specific N:P that would cause a shift from, for instance, P to N limitation. Such shifts could simply be the results that human inputs were greater for one nutrient, which could ultimately result in limitation of the other. These variable limitation patterns could also be the consequence of among stream differences in: algae composition (Borchardt 1996), different site-specific physical conditions (Rosemond et al. 2000), or the nature of human-caused nutrient inputs (i.e., pulses vs press)(Portielje and Lijklema 1996). Temporal variation may also play a role. Streams are known to shift limitation from one nutrient to another seasonally, but whether or not this occurs depends on the extent to which the relative abundance of the predominant algae taxa changes (Hullar and Vegal 1989). While our study was limited to a single season, the extent to which algae composition changes seasonally differs among streams, which could lead to variable limitation patterns among all of these streams.

There are considerable management implications to human caused shifts in patterns of nutrient limitation because the relative importance of N vs. P may change along recovery trajectories. Under such circumstances, controls for both N and P are more likely to improve ecological responses. Another implication for enriched streams, relates to the nutrient augmentations intrinsic to NDS experiments. A better experiment might be one that estimates limitation patterns under nutrients removal scenarios, but this may be difficult to accomplish *in situ*. Another possibility would be evaluating changes in C:P and C:N in algae tissue following reciprocal transplants of cobbles from high and low nutrient streams. King et al. (2000) tested this possibility under experimental nutrient additions in artificial stream channels and found that the C:P of algae from low nutrient streams declined with increasing nutrient treatments, whereas the C:P of algae assemblage from high nutrient streams did not change (i.e., the algae were not P-limited).

Our decision to define enrichment classes based on the primary sources of nutrients resulted in a site on the Weber River that was not enriched in comparison with other highly enriched streams. This site was among the largest in our study with base flows of ~35 cfs) and also had one of the smaller POTW discharges (0.8-1.4 cfs). In this case, the small discharge did not increase downstream nutrient concentrations (Table 3.2). The comparatively low nutrient concentrations at this site explain why it behaves differently, in terms of limitation, than the majority of sites in the highly enriched class. These results highlight the importance of not making generalizations of anticipated nutrient responses based exclusively on sources (i.e., comparisons between point source and nonpoint sources).

Saturation Thresholds

We found thresholds for both TN and TP that define the ambient conditions associated with saturated conditions. These thresholds are certainly not protective, but they do help provide an upper benchmark to the other indicators that we evaluated. An ecologically important implication of these benchmarks relates to the relative assimilative capacity of streams. As nutrient concentrations within streams increases the uptake velocity decreases (Earl et al. 2006, Obien and Dodds 2010). This means that as streams near saturation thresholds they are increasingly unable to perform the ecosystem service of nutrient retention. Further incremental increases in nutrients at such sites would be transported further downstream leading to an expansion in the spatial scale of nutrient-related impairments. Other investigations have also observed nutrient saturation associated with patterns of land use. Bernot et al. 1996 observed N to be at saturation and P to be near saturation at watersheds with high levels of agriculture. Others have observed similar conditions along gradients of increasing urbanization (Meyers et al. 2007).

We tested the significance of our saturation thresholds using ROC analysis and then by comparing limitation among all sites above and below the TN and TP thresholds. The results of ROC suggest that our thresholds predict the presence or absence of NDS nutrient limitation quite well. We found that for TN the true positive (sensitivity) and true negative (specificity) prediction rates were reasonably balanced (0.78 and 0.75, respectively). For TP we found a different condition where sensitivity was fairly low (0.56) but specificity was high (0.80). This indicates that above the TP threshold of 0.078 mg/L there are nearly even odds that a site will not be nutrient limited, whereas below this threshold we have higher confidence that sites will be nutrient limited as the threshold predicts.

In general, saturation thresholds patterns suggest that, on average, N may be slightly more important than P in controlling algae accrual. For instance, the ROC models indicate that TN saturation thresholds are stronger predictors of nutrient limitation than TP. Similarly, sites where ambient nutrients were below both the TN and TP thresholds algae were primarily N and secondarily P limited. However, as emphasized below, one should be cautious about concluding from these general patterns that P reductions are not needed at streams with eutrophication problems. Again, these are general patterns, whereas remediation strategies are intrinsically site-specific.

Management Implications

Regional NDS studies, such as this one, provide several useful measures of nutrient responses. For instance, these studies identified in-stream concentrations of TN and TP that, on average, are likely to be high enough to saturate NDS algal growth. Among other things, saturation concentrations are important in the context of nutrient remediation efforts because one would not expect to see improvements in algae growth until ambient nutrient concentrations fall below saturation thresholds. Moreover, once the nutrient assimilative capacity of streams is exceeded further nutrient inputs have a greater potential of degrading downstream uses.

NDS bioassays can also help inform and prioritize nutrient reduction efforts. For instance, by understanding which nutrient limits algal growth in a system, managers can focus their resources on reducing the nutrient that will have the greatest improvement on downstream water quality, instead of implementing a "one-size-fits all" nutrient reduction strategy. However, resource managers should be cautious over interpreting NDS experimental results. While we were able to identify several general patterns of nutrient limitation, there were always exceptions. Overall, these results highlight the importance of addressing, to the greatest extent possible, both N and P when implementing nutrient reduction strategies. For decades, the assumption that P limits primary production in freshwaters was a paradigm in aquatic ecology. Indeed, UDWQ historically established TMDL limits exclusively for P to address nutrient-related water quality concerns. Our results suggest that N limits should also be included, or at the very least considered, in development of NNC and associated TMDLs. The fact that the addition of both N and P frequently exhibited a greater response than either nutrient alone suggests that the simultaneous reduction of both N and P may be the most effective remediation strategy. Moreover, the differences that we observed in nutrient limitation patterns among enrichment classes suggests that one might expect to see shifts from N to P, and vice versa, once best management practices (BMPs) are implemented.

Water resource managers rely heavily on water chemistry samples as the backbone of their regulatory programs. Stream nutrient concentration thresholds above which algal growth is likely to be unlimited by nutrients (i.e., saturated) provides valuable information for stream assessment and resource prioritization. Such thresholds may prompt further studies needed to evaluate the need and efficacy for nutrient reduction efforts. N or P saturation thresholds of 0.42 mg/L TN and 0.078 mg/L TP will be used as one indicator, among others, to inform NNC development for headwaters or in an assessment context to help identify nutrient concentrations of potential concern, which provides additional context to data already routinely collected by UDWQ and collaborators.

Overall, we conclude that NDSs provide meaningful measures of stream functional responses. Moreover, because these experiments are inexpensive, these experiments will be used in the future, as needed on a site-specific basis. However, management decisions based upon NDS data must also consider other site-specific observations and nutrient responses. For instance, excessive benthic algal growth is not likely to be the most important cause of degraded uses in low gradient, soft-bedded streams, nor in larger rivers where benthic algae growth is light-limited. Moreover, algae-bacterial production may become increasingly decoupled in high nutrient streams (Scott et al. 2008), which suggests that the relative importance of autotrophic or heterotrophic nutrient limitation may also differ among different types of streams. In cobble-bedded streams, where excessive benthic algae growth is more likely to be an important nutrient response, other factors may also be important and immediate stressors to stream biota. For instance, in streams with low nighttime DO concentrations, heterotrophic responses to nutrient inputs may be a more important factor in the degradation of aquatic life uses. Streams that are slow-moving and depositional in nature are more complicated because they are inherently characterized by large changes in habitat, particularly increases in the number of depositional zones. This results in accrual of unstable smaller sediment particle size and increased deposition of both allochthonous and autochthonous organic debris. By nature, such stream habitats support different biota than those with greater velocity and fewer depositional zones. Despite these limitations, NDS experiments certainly have an important place in the "tool box" of techniques that can be used quantify ecological responses to nutrient enrichment.

STREAM METABOLISM

Introduction

Whole stream metabolism is a measurement of ecosystem function that includes ecosystem-scale rates of photosynthesis (gross primary production, GPP) and respiration (ecosystem respiration, ER). The relative rates of GPP and ER in an ecosystem identify the basal source of energy supporting the aquatic food web: allochthonous (from outside the system) or autochthonous (produced within the system). Stream metabolism, and the calculation thereof, is based on the premise that changes in dissolved oxygen (DO) concentrations—between daytime highs to nighttime lows—are the result of photosynthesis (biologic production of O_2), respiration (biologic consumption of O_2), and reaeration (bidirectional atmospheric exchange (Figure 4.1).

Researchers have used stream metabolism to investigate rates of GPP and ER since the pioneering work of Odum (1956). Since that time stream metabolism has primarily been investigated

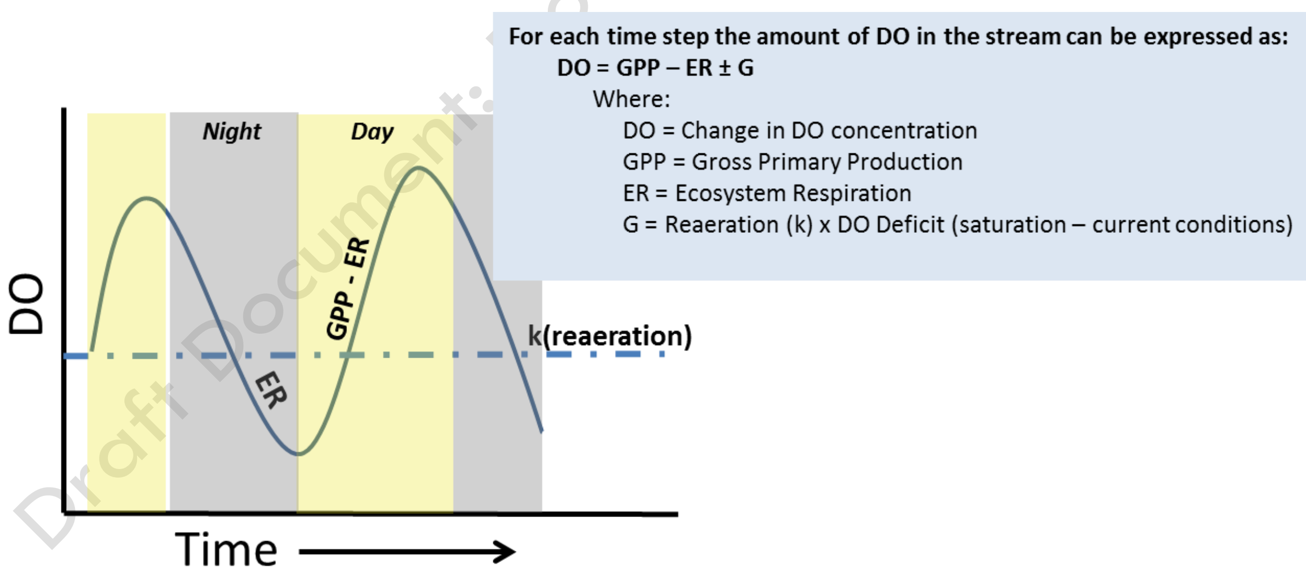


Figure 4.1. Conceptual model that depicts how Ecosystem Respiration (ER) and Gross Primary Production (GPP) relates to daily fluctuations in Dissolved Oxygen (DO) concentrations. As the sun set, GPP rates decline because plants and algae have insufficient light to photosynthesize. Ultimately, ER consumption is greater than GPP so DO starts to decline. At nighttime, GPP is zero, so the amount of DO in the stream is a function of the physical process of reaeration (k), which adds water to the stream and the respiration of all organism in the stream ecosystem. As the sum rises, GPP will start increasing DO, some of which continues to be consumed by stream organisms (ER).

using two techniques: mesocosm (bottles or chambers) experiments and *in situ*—or "open channel"—methods. Recently, open channel techniques have gained widespread acceptance for several reasons. First, underlying data required for metabolism calculations is now reasonably accessible due to the availability of high quality, relatively low cost DO sensors and data loggers. Second, open channel methods more accurately reflect reach-scale conditions because they do not introduce container effects. Finally, these techniques also avoid scaling problems that can arise as one extrapolates mesocosm results to stream reaches, which is the scale of interest to resource managers. Indeed, open channel metabolism is sometimes call "whole stream" metabolism because it integrates all the metabolic processes and surface water-benthos interactions that occur over an entire stream reach (Young et al. 2008, Izagirre et al. 2008).

Aquatic ecologists have investigated both natural and anthropogenic landscape-scale controls on whole stream metabolism such as geography (Hill et al. 2000, Bernot et al. 2010), land use practices (Young and Huryn 1999, Houser et al. 2005), and riparian disturbance (McTammany et al. 2007). Others have investigated how stream metabolic rates influence ecological processes, such as nutrient processing (Hall and Tank 2003) and ecosystem structure (Sabater et al. 2002). Together, these studies, among others, show whole stream metabolism has the potential to be an excellent indicator of stream condition. Both GPP and ER integrate several reach-scale factors that influence stream health: geomorphology, hydrology, riparian vegetation, in-stream vegetation, climate, biology, and chemistry (Mulholland et al. 2005, Grace and Imberger 2006, Young et al. 2008). On the other hand, many of these same factors vary naturally, so condition measures will need to decouple natural versus human-caused variation for metabolism to be useful for assessing specific sites.

We obtained whole stream metabolism for 31 streams along a gradient of ambient nutrient concentrations to evaluate the potential use of stream metabolism as a functional indicator of nutrient enrichment. Accordingly, we compared daily rates of GPP and ER to stream nutrient concentrations. We also evaluated the extent to which these relationships were influenced by several potentially important covariates (e.g., stream slope, shading and turbidity). We then evaluated ties to existing regulation by evaluating the extent to which daily rates of GPP and ER were associated with Utah's DO water quality criteria. Finally, we developed multivariate models to evaluate the extent to potential covariates influence GPP and ER estimates and then discuss how the influence of important covariates could be reflected in resource management decisions.

Methods

Data Collection

We obtained DO measurements for measured whole stream metabolism measurements at 34 sites (see Figure 2.1, Chapter 2), although 6 sites were subsequently excluded from further metabolism analysis due to their extremely high turbidity (details below). This left us with a total of 31 sites about half of which (15 sites) were in reference condition. At each site, we deployed a water quality probe (YSI 6600V2 or 600 OMS V2) to measure dissolved oxygen (DO; see Appendix A for the SOPs) and temperature at five-minute intervals for a minimum of 48 hours. We obtained solar radiation data from the closest available weather station (mesowest.utah.edu). Surface water nutrients were collected at the time of sonde deployment and retrieval and were analyzed for total nitrogen (TN) and phosphorus (TP) at the Aquatic Biogeochemistry Laboratory at Utah State University (Valderamma 1981).

Construction of Metabolism Models

We calculated stream metabolism using an open water method with reaeration (K) as a free parameter (Hall et al. 2014) based on the following equation derived from Van de Bogert et al. (2007):

$$O_t = O_{t-1} + \left(\frac{GPP \cdot \Delta t}{z} \times \frac{Light_t}{\sum Light} + \frac{ER \cdot \Delta t}{z} + K(O_{sat} - O_{t-1}) \cdot \Delta t\right)$$

Where,

ER = Ecosystem Respiration (loss of g O^2/m^2/day)
GPP = Gross Primary Production (g O^2/m^2/day)
K = Reaeration coefficient (day^{-1})
Light = Solar radiation or PAR
O = Dissolved Oxygen (mg/L)
O_{sat} = Oxygen Saturation (mg/L)
t = Time (fraction of day)
z = Mean stream depth (m)

This metabolism model adjusts GPP and ER at each time step to fit the oxygen data using non-linear minimization (R function `nlm`) of the maximum likelihood accuracy estimates. In this equation, K can be modeled as a free parameter from the oxygen data simultaneously with GPP and ER. In rare cases where K could not be modeled accurately, we had to constrain K with values calculated from nighttime regression (Grace & Imberger 2006) to improve model performance.

Comparison of GPP and ER to Stream Nutrient Concentrations

We used linear regression to evaluate the relationship between nutrients (TN and TP) and the metabolic response rates GPP and ER. To further explore the relationships among nutrients and metabolism metrics we classified sites into three groups with similar TN concentrations, then three groups with similar concentrations of TP (hereafter nutrient groups). We opted for three groups to be consistent with UDWQ and USEPA assessment methods, which generally assess waterbodies into three condition classes (UDWQ 2014). In our case, high medium and low nutrient groups were empirically derived using deviance reduction with bootstrapping. NDR is a nonparametric procedure that uses least squares fitting to identify, from all possible splits, groups where among stream differences in nutrient concentrations are as homogeneous as possible within groups, while also maximizing among group dissimilarity. In this case, quantitative estimates of group similarity—and dissimilarity —were based on similarity in the mean and variance of nutrient observations at each site (Qian et al. 2003, NDR, package `rpart`). We subsequently used ANOVAs followed by post-hoc Tukey's Honestly Significant Difference (HSD) to determine if there were significant differences ($p<0.05$) in daily rates of GPP and ER among the three nutrient groups. Our assumption with this analysis was that if daily rates of GPP and ER significantly differed among the three nutrient groups, then corresponding thresholds associated with the metabolism metrics could be used to determine the relative degree of nutrient enrichment among streams.

Derivation of GPP and ER Indicators

We used identical NDR procedures to define condition classes from GPP and ER responses. As with the nutrient groups we opted to establish three groups that define good, fair and poor GPP and ER conditions. The metabolic thresholds established from these procedures were used for DO comparisons and will ultimately be used to identify sites with nutrient-related problems.

Comparisons to DO Numeric Criteria

Our estimates of GPP and ER thresholds are intrinsically statistical and do not necessarily translate to the health stream ecosystems. To provide a more direct linkage, we evaluated whether metabolic rates corresponded to more traditional interpretations of DO as a water quality parameters, to validate metabolism as a functional indicator of nutrient enrichment. For these first pass validation exercises we compared the DO data used to make the metabolism calculations against the independently established numeric DO criteria for the most sensitive use of each site (Table 4.2). We made these comparisons using two of Utah's DO criteria—the daily minimum when early life stages are not present and 30-day average—to represent acute and chronic threats to stream biota (UAC R317-2-14, Table 4.1). We then compared the extent of oxygen criteria violations among the three GPP and ER groups using ANOVA ($p<0.05$) and then a *post-hoc* Tukey's HSD test.

Evaluation of Potential Covariates

Rates of GPP and ER vary naturally, so we also evalated the relative influence of nutrients and other natural environmental gradients. We used multivariate Random Forests (Breiman 2001, R package `randomForest`) to model ER and GPP from 20 potential explanatory variables that

Table 4.1 Results of Random Forest models that were used to explore the influence of 20 candidate covariables on Gross Primary Production (GPP) and Ecosystem Respiration (ER). Importance of variables was evaluated using the % increase Mean Squared Error (MSE). Higher MSE indicates that when values in a variable were randomized the model performance declined. Data were obtained from the Utah State University Aquatic Biogeochemistry Laboratory (USU ABL), Utah Unified Public Health laboratories (UPHL), U.S. Geological Survey Stream Stats program (USGS) or the Utah Division of Water Quality Comprehensive Assessment of Stream Ecosystems program (UDWQ).

	Units	GPP %Increase MSE	ER %Increase MSE	Source
Total Nitrogen	mg/l	72.6	69.0	USU ABL
Total Phosphorus	mg/l	73.9	52.1	USU ABL
Turbidity	NTU	41.7	37.0	UPHL
Total Suspended Solids	mg/l	50.0	35.2	UPHL
Channel Shading	%	104.3	69.3	UDWQ
Slope	%	103.1	63.4	USGS
Basin Area	mi^2	58.1	30.4	USGS
Herbaceous Upland	%	69.1	25.7	USGS
Forested Watershed	%	67.5	32.1	USGS
Basin Slope	%	75.3	42.3	USGS
Mean Water Depth	cm	52.6	52.6	UDWQ
Mean Thalweg Depth	cm	32.7	28.0	UDWQ
Bankfull Hieght	cm	69.4	47.3	UDWQ
Channel Incised Hieght	cm	58.6	13.7	UDWQ
Channel Width:Depth Ratio		19.2	30.5	UDWQ
Fine Substrate (<2mm)	%	41.2	19.0	UDWQ
Small Sediment (<16mm)	%	64.5	41.9	UDWQ
Median Particle Size	cm	71.6	46.9	UDWQ
Riffle and Rapid Channel Units	%	38.2	32.5	UDWQ
Riparian Corridor Bare Ground	%	28.0	25.7	UDWQ

capture, directly or indirectly, characteristics that are known to—or have been suggested to— control whole stream metabolism. Potential explanatory variables were obtained from GIS (USGS StreamStats), on-site physical habitat surveys (USEPA 2009), and water quality samples (Table 4.1). In some cases, we opted to use general landscape-level surrogates of stream size (e.g., basin area, water depth, width:depth) as opposed to direct measures of physical characteristics (i.e., temperature and substrate composition) for several reasons. First, these stream characteristics generally vary systematically from headwaters to larger streams, so these landscape-scale characteristics simultaneously capture several important stream characteristics. In addition, landscape-level characteristics integrate steam characteristics, both spatially and temporally, at scales that ae more closely aligned with the scale of these regional analyses. For instance at most sites we only had about three days of temperature data and whatever summary statistic of temperature we selected for data over this range could never capture ecologically relevant among stream differences is the thermal regimes. Another consideration was pragmatic with respect to management applications for this work. The integrative nature of these landscape-scale characteristics potentially provides more parsimonious relationships that could be more readily applied on a statewide scale.

We first ran Random Forest regression on all variables and then selected the best performing variables—based on percent increase of <u>m</u>ean <u>s</u>quare model <u>e</u>rror (MSE) that occurred under model runs where one-by-one the observations for each variable were randomly assigned to another stream while other variables remained constant. The underlying assumption MSE variable selection is that the most important variables will result in the largest decrease in model accuracy when the observations for that variable are randomly reassigned. The magnitude of change in model performance provides an estimate of variable strength that is internally consistent with the model. After selecting the most important factors that influenced GPP and ER, we re-ran the analyses to create final models that reflected a reasonable balance between model accuracy and parsimony. For instance, if the best subset of variables performed as well as the all variables Random Forest model (based on the pseudo-r^2 fitness statistic) then we considered the best subset model successful. The goal of this exercise was to identify important covariates that could potentially obscure or exaggerate the role of that nutrients play in determining GPP or ER rates. All analyses were conducted in R v2.15.0 (R Core Development Team, 2012).

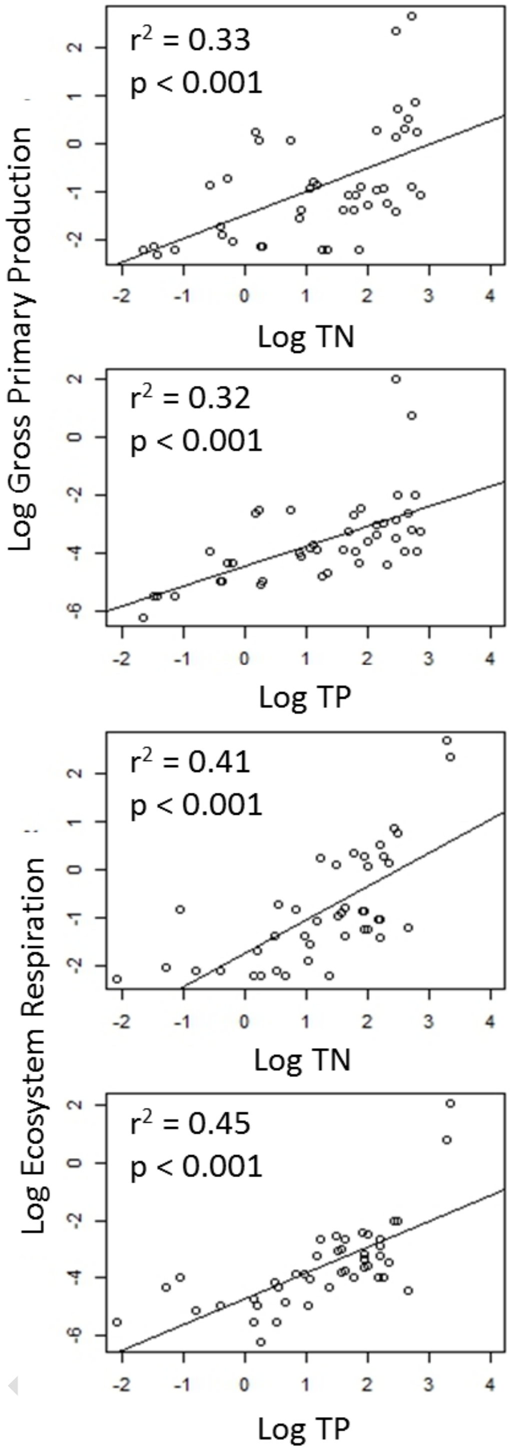


Figure 4.2 Gross Primary Production and ecosystem respiration as a function of Total Nitrogen and Total Phosphorous.

Results

Early exploratory analyses revealed metabolism rates were suppressed at highly turbid sites as were relationships between nutrients and rates of GPP and ER. Distributions of turbidity data revealed five highly turbid outliers with a turbidity of greater than 75 Nephelometric Turbidity Units (NTU). These five highly turbid streams were subsequently excluded from the remainder of the analyses. Nevertheless, the remaining 31 streams still encompass a broad nutrient gradient (TN 0.10-14.37 mg/L and TP 0.002-7.65 mg/L) and should not overly bias the remaining analyses.

Relationships between Metabolism and Nutrients

Simple linear regressions revealed significant relationship between nutrients (TN and TP) and functional responses (GPP and ER) across all non-turbid sites (Figure 4.2). GPP was positively related to both TN (r^2 = 0.303, p<0.001, Figure 4.2A) and TP (r^2 = 0.372, p<0.001). Correlations between ER and nutrients were slightly stronger, for both TN (r^2 = 0.471, p<0.001, Figure 4.1C) and TP (r^2 = 0.485, p<0.001, Figure 4.1).

From the deviance reduction models we identified three distinct groups of streams that were similar with respect to the mean and variance of TN and TP observations (hereafter Low, Medium and High nutrient groups, Table 4.2). These statistical analyses identified the same thresholds for GPP and

ER. TN values of 0.24 mg/L and 1.28 mg/L, and TP values of 0.02 mg/L and 0.09 mg/L separate low medium and high rates of both GPP and ER.

These TN and TP nutrient groups generally corresponded predictably with measures of stream metabolism. However, post hoc investigations revealed that among all streams, all three nutrient groups were only statistically distinct for comparisons of ER with TN (Figure 4.3). For all other relationships the among stream metabolism metrics were only able to distinguish the low nutrient group from all others. Among all sites GPP and ER rates differed among the three nutrient groups for both TN (ANOVA, $p<0.001$) and TP (ANOVA, $p<0.001$) (Figure 4.2).

GPP differed predictably among nutrient groups with higher rates generally corresponding to sites with higher nutrients. For TN, among group mean GPP rates (g O^2/m^2/day) ranged from 2.43±3.27 (standard deviation) for the low N group to 6.57±4.9 for the medium group, and then 13.19±2.59 for the high group. GPP rates were similar for the TP nutrient groups: Low=3.62±4.74, Medium =7.48±4.75, and High=13.86±2.29) (Figure 4.2).

ER also differed predictably among TN groups. From low to high, mean ER rates (g O^2/m^2/day) among the TN groups ranged from 2.05±2.28, to 5.78±3.29 and then14.35±9.35. Mean ER rates for the low, medium and high TP groups were 3.13± 3.81, 6.05±2.31, and 19.66±9.25 respectively (Figure 4.3). For these streams, group mean ER rates were generally not dependent on whether the groups were established from TN or TP, although respiration may be slightly higher at sites with the highest TP concentrations in comparison with the high TN group. For both nutrients, ER rates were similar to those observed for GPP for the low and medium groups, whereas ER was consistently higher than GPP at sites with the highest nutrient concentrations.

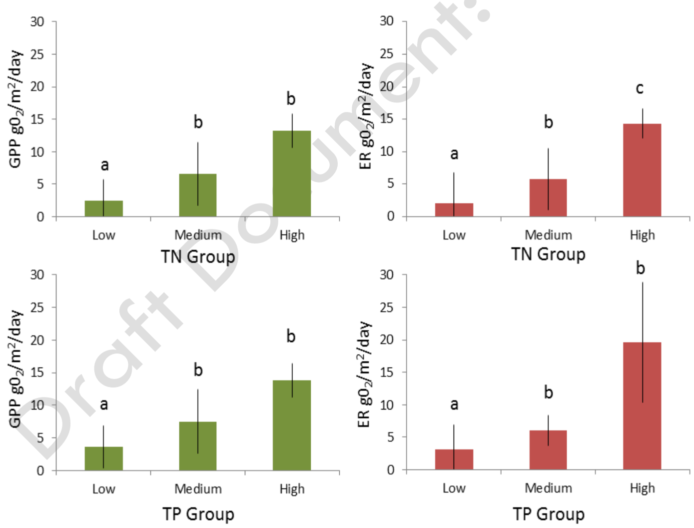


Figure 4.3. Bar chart comparing daily rates of GPP (green) and ER (red) among low, medium and high concentration sites for TN and TP. Specific group thresholds are shown in Table 4.2. Letters above bars indicate significant differences (Tukey's HSD). $p<0.05$)

Table 4.2. Empirically derived thresholds for nutrients and the functional responses GPP and ER. Nutrient and metabolism thresholds were derived independently to divide sites into good, fair or poor condition classed.

Nutrient	Nutrient Group Thresholds	Functional Indicator	Indicator Group Thresholds
TN (mg/l)	Low < 0.24 > Medium < 1.28 > High	GPP (g O_2/m^2/day)	Good < 6.0 > Fair < 10.0 > Poor
TP (mg/l)	Low < 0.02 > Medium < 0.09 > High	CR (g O_2/m^2/day)	Good < 5.0 > Fair < 9.0 > Poor

Stream Metabolism Groups

NDR thresholds were established independently for GPP and ER to establish good, fair and poor condition classes for each metric (Table 4.2). For GPP, we identified a threshold of 6 g O^2/m^2/day to distinguish between good and fair condition classes and another threshold of 10 g O^2/m^2/day. The two thresholds for ER were a little lower at 5 g O^2/m^2/day and 9 g O^2/m^2/day. All groups had statistically significant (ANOVA, $p<0.05$) differences in metabolic rates.

Relationship among Metabolism Metrics and DO Criteria

We also evaluated the extent to which the metabolic condition classes were associated with excursions below several DO criteria with varying averaging periods, because these are independently derived indicators of potential threats to stream biota. The minimum daily DO concentration was significantly different among the three nutrient groups for both GPP and ER (ANOVA, GPP $p<0.001$ and ER $p<0.001$). The absolute minimum DO observed during sonde deployments was generally higher at sites in the "good" metabolic groups.

The metabolism functional responses also corresponded predictably to Utah's DO criteria for both absolute minimum (acute) and 30-day average (chronic) averaging periods. We found significant differences among GPP and ER groups with the relative frequency that samples at each site fell below minimum DO water quality criteria (ANOVA, GPP $p<0.001$ and ER $p=0.018$). For GPP, we found that samples fell below acute and chronic DO criteria less often for sites in the "good" metabolic groups (Tukey's HSD, $p<0.05$) than for sites in the "fair" or "poor" groups. The same tests revealed similar patterns among the ER groups. On average, DO observations at sites in the poor GPP and ER class exceeded acute (minimum) DO criterion ~6% of the time and chronic (30-day) criterion ~ 45% of the time. Of course, these general trends obscure important site-specific differences. The Fair and Poor GPP and ER groups had many sites that fell below the DO criteria and many sites that did not, which led to large within-group variation in minimum DO criteria violations (Figure 4.3).

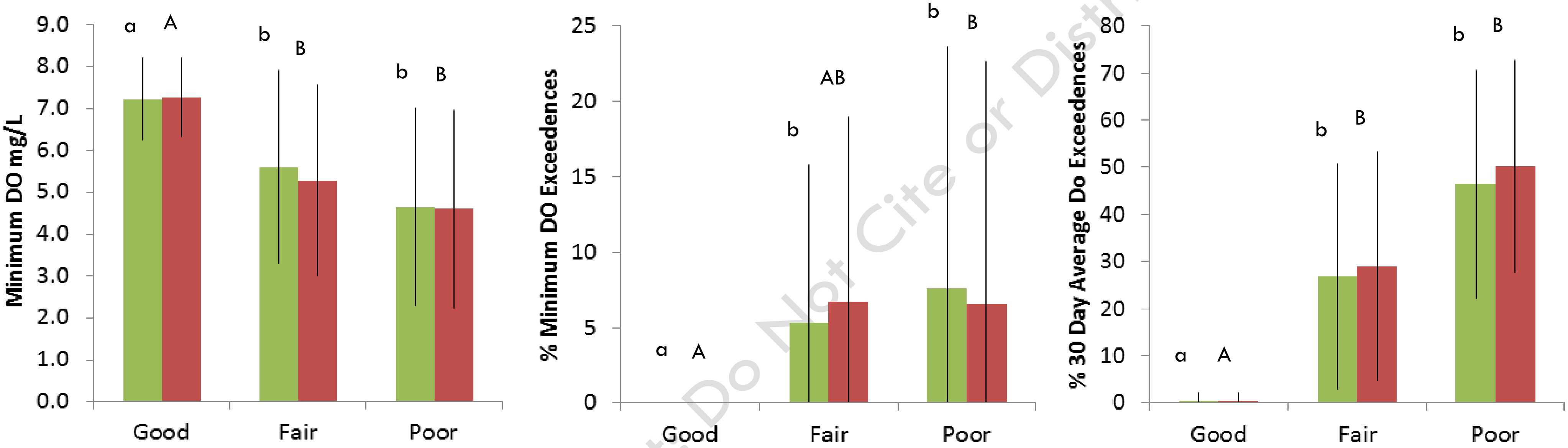


Figure 4.4. Comparisons of three measure of oxygen dynamics with three groups of streams with varying daily GPP (green bars) and ER (red bars) rates (Good, Fair, Poor, Table 2). Lower case letters indicate significant differences of GPP groups and upper case indicates significant difference of ER groups determined by an ANOVA and *post-Hoc* Tukey's HSD. Error bars are standard deviation.

This page intentionally left blank

Physical Covariates

We ran Random Forest regression models separately to GPP and ER from nutrients and 20 potential covariates (Table 4.1) obtained from water quality samples, GIS analyses and site-specific habitat metrics. With all variables our models found significant relationships with GPP (mean squared residuals = 12.7, pseudo r^2=0.54) and ER (mean squared residuals=18.3, pseudo r^2=0.45). Four of the top five predictor variables were the same for GPP and ER, as measured by increases of mean square error (MSE). The top predictor variables for GPP were stream slope (MSE=103.9), stream shading (103.3), basin slope (74.9), TN (73.4), and TP (72.6). Similar variables were most important for predicting ER, including shading (MSE=70.2), TN (68.9), stream slope (63.2), mean stream depth (53.8), and TP (51.4). We ran Random Forest regression again with only the top four variables that were found in GPP and ER to compare overall model performance (stream slope, shading, TN and TP). We found that the model performed just as well with only the top four variables for GPP (mean squared residuals = 13.1, pseudo r^2=0.53) and ER (mean squared residuals = 16.2, pseudo r^2=0.51).

We explored relationships with channel slope and shading further to identify thresholds that could potentially be used to modify ER and GPP expectations, potentially increasing the accuracy of metabolism assessments. NDR revealed significant thresholds at ~1% slope for both ER and GPP. GPP and ER thresholds were also found for percent channel shading, where streams with channel shading less than ~11% had greater mean daily rates of GPP (9.3±5.6 to 3.99±4.1) and ER (8.10±5.5 to 4.31±4.1). While these two variables alone cannot account for all of the variation in GPP and CR, they will undoubtedly improve the interpretation of metabolism responses.

Discussion

Nutrient Thresholds

Using daily rates of GPP and ER we found two thresholds of TN and TP that can be used to demark concentrations where nutrient enrichment generally alters stream metabolic functions (Table 4.2). TN values of 0.24 mg/L and 1.28 mg/L and TP values of 0.02 mg/L and 0.09 mg/L separate low, medium, and high rates of both GPP and ER. UDWQ and collaborators can use these thresholds in combination with those obtained from other functional and structural indicators to identify nutrient concentrations where nutrients are sufficiently high to affect stream conditions.

Comparison to Numeric DO Criteria

One way excess nutrients cause deleterious effects to stream biota is through alteration of diel oxygen dynamics via increased autotrophic or heterotrophic productivity. Stream metabolism provides ideal metrics to evaluate those effects because it directly quantifies the biological processes responsible for alterations to DO dynamics. By statistically binning daily rates of metabolism into three categories we were able to demonstrate significant differences among the absolute minimum DO observed at each site and percent of times that DO observations were lower than minimum DO criteria (Figure 4.3), which provides evidence that these conditions can be directly tied to independent measures of designated use support.

UDWQ established DO criteria to protect aquatic life for each of three beneficial uses: coldwater fisheries (3A), warmwater fisheries (3B), and non-game fish fisheries (3C) (UAC R317-2-6). Each of these beneficial uses has a different minimum DO criterion based on differing sensitivity of fish, and other organisms in their food web. Minimum DO water quality criteria quantify, given assumptions intrinsic with DO standards development methods, conditions where short-term exposure is of potential threat to stream biota. Among all sites in this study, only 11% had a violation of the applicable minimum DO criterion. None of these sites had >10% of the observations below this criterion, which is the exceedance frequency that UDWQ currently uses for assessments purposes. UDWQ is currently reviewing whether the"10% rule" is appropriate for continuous data. Also, these data may underestimate the number of sites with an extremely low DO concentration, because our data were collected in the summer and others have found that acute anoxic conditions occur in the autumn following algae senescence (Suplee, 2012). On the other hand, the extent to which these low DO values actually threaten fish may need further investigation because fish are highly motile and may be able to find microhabitats that limit their exposure to low DO conditions, particularly if the conditions are relatively short lived.

We also explored exposure to chronic DO conditions by comparing the percent of daily minimum DO observations for each site that fell below the appropriate 30-day average DO criterion assigned to each site and found that sites in the poorest condition, as measured by metabolism metrics, exceeded this criterion 45% of the time. We acknowledge that these short-term observations (48-72 hour) are not representative of actual 30-day averages and that these estimates may over- or underestimate chronic DO exposure. Nevertheless, this analysis demonstrates clear connections between GPP and ER response metrics and designated use support. In fact, UDWQ currently uses the 30-day average for assessment purposes because we assume that this value is more reflective of long-term conditions. UDWQ is currently evaluating alternative methods for using instantaneous DO measurements for assessment purposes. At a minimum, these data suggest that sites with atypically high rates of summertime GPP and ER warrant follow-up investigation to determine whether or not low DO is also a concern.

GPP and ER Thresholds

Our thresholds of stream condition for GPP (6.0 and 10.0 $gO_2/m^2/day$) and ER (5.0 and 9.0 $gO_2/m^2/day$) are similar to the suggested rates of GPP and ER proposed by Young et al. (2008) as indicators of river health in New Zealand rivers (7.0 and 9.5 $gO_2/m^2/day$ GPP and ER, respectively). The independently derived New Zealand metrics were obtained from a meta-analysis of metabolism calculations obtained from numerous reference sites over a 16-year period of record. Our stressor-response approach, along with Young's reference condition approach, are part of the growing literature that provides general guidelines about the GPP and ER rates that are reflective of healthy conditions. However, the results of our Random Forest models also highlight the importance of taking natural changes in stream conditions into account before universally applying thresholds to infer stream condition.

Physical Covariates

We found that nutrients were unrelated to metabolic rates at sites where turbidity was greater than 75 NTUs, which likely stems from a lack of light reaching autotrophic benthos. We suggest that stream metabolism is not an appropriate functional indicator for these sites. Nevertheless, at these highly-turbid sites the mean TN (2.41 mg/L) and TP (0.36 mg/L) concentrations were an order of magnitude greater than the highest NNC proposed elsewhere, so other indicators are likely to detect nutrient related impairments at these sites. Moreover, these streams are atypical because streams with >75 NTU comprise less than three percent of the total stream miles in Utah (UDWQ, unpublished data). Nonetheless, these data also highlight the importance of understanding the relative influence of multiple stressors on aquatic life degradation, because excess sedimentation at these sites may be a more immediate threat to stream biota at these sites, despite high nutrient concentrations. In arid environments like Utah some streams are naturally turbid and may be less susceptible to the deleterious effects of nutrient enrichment. Stream metabolism metrics could provide a way of documenting that GPP and ER remain protective in highly turbid streams.

Rates of both GPP and ER are also influenced by natural conditions. For instance, sites with channel slopes <1% had higher rates of GPP and ER than those above the threshold. Although this relationship cannot be entirely attributed to natural conditions because slope was also strongly related to both TN (Pearson correlation r=-0.603) and TP (r=-0.617, data not shown). This relationship is not surprising considering that anthropogenic nutrient sources, including agricultural activities and urban discharges, are more likely to be concentrated at lower gradient stream segments, where most of Utah's population resides.

Among all of the covariates that we evaluated, channel shading has the clearest influence on stream metabolism of all the parameters measured, the rates of both declining with increasing channel shading. Channel shading was also only weakly correlated to TN (Pearson correlation r=-0.245) and TP(r=-0.221, data not shown). Both nutrients and shading are important determinants of stream metabolism, but the effects of these factors are very different and frequently unrelated. Nutrients elevate GPP and ER, whereas shading represses GPP. Moreover, as the poor correlations illustrate, streams with high channel shading can occur in both streams with both high and low nutrients. The potential effect of shading on GPP is straightforward, but the effect on ER is less intuitive and a topic of debate among stream ecologists. As we observed, other region-scale estimates of stream metabolism have noted a close correspondence between ER and GPP (Benot et al. 2010). Others have observed that ER rates are tied to rate of bacterial (heterotophic) production (Murray et al. 1987). Rier and Stephenson (2001) found that algal cell biomass was the best predictor of bacterial cell density unless benthic chlorophyll concentrations were low, which primarily occurred in low nutrient streams. These observations suggest that competition for nutrients between autotrophs and heterotrophs, particularly in streams where GPP is low. However, other demonstrated a relationship between autotrophs and heteroptrophs in biofilm, yet were unable to document competition among these assemblages (Carr et al. 2005). One possible explanation is that the mechanism behind algal and bacterial relationships depends on the extent of nutrient enrichment. This possibility is supported by the fact that in high nutrient streams algal-bacteria production become decoupled, whereas they remained closely coupled in low nutrient environments (Scott et al. 2008).

Our results on the influence of channel slope and shading to GPP and ER are in agreement with a number of stream metabolism studies. For instance, high flow events, which are often more extreme at higher slope sites or highly urban sites, reduce daily rates of GPP by increasing stream scour, and ER mainly by exporting accumulated organic matter (Uehlinger et al. 2003, Acuna et al. 2004). Streams with high slope likely accumulate less organic matter. Hill and Dimick (2002) quantified the importance of shading relative to other factors in determining GPP rates by artificially manipulating stream shading and then quantifying the corresponding decline in photosynthesis rates. Bott et al. (2006) and Bernot et al (2010) found that photosynthetic radiation (PAR) better explained among stream differences in GPP than any measurement of watershed land use. These studies indicate that even under heavily modified land uses (agriculture or urban) GPP can remain relatively low if a healthy riparian corridor is maintained. Or conversely, in ecosystems with little natural riparian vegetation GPP would be highly responsive to anthropogenic increases in nutrients. Together, these studies suggest that threats of low DO can be minimized by protecting, or restoring, the riparian corridor. They also highlight the importance of considering multiple stressors in Utah's Nutrient Reduction Strategy.

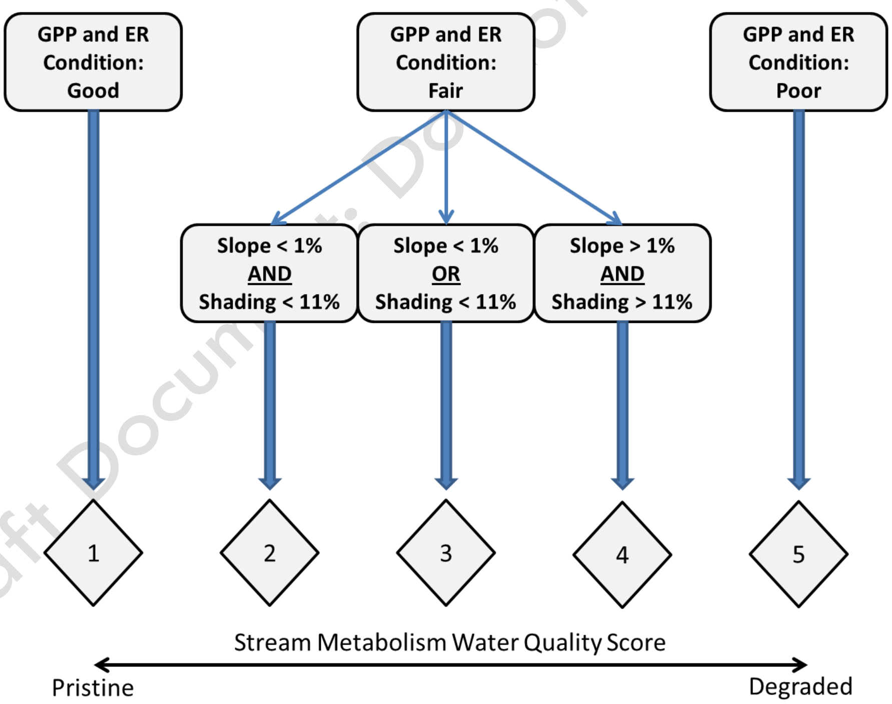


Figure 4.5. A hypothetical model of how important covariates might be incorporated into interpretations of stream metabolism responses metrics in an assessment context. For instance, high GPP or ER observations at high gradient streams with an extensive canopy would provide more conclusive inference of water quality problems than other streams where physical conditions naturally cause increases in these metabolism metrics.

We developed a framework to facilitate the interpretation of GPP and ER and compensate for key covariates (Figure 4.5). We suggest using slope and shading to evaluate sites that fall within the Fair range of GPP (6-10 g O^2/m^2/day) and ER (loss of 5-9 g O^2/m^2/day) rates. GPP and ER are naturally lower at sites with high slope (>1%) or high channel shading (>11%). If we observe high GPP or ER rates at sites that meet both of these conditions, then we can be fairly confident that that the metabolic rates have been elevated by eutrophic conditions. At a minimum, elevated GPP or ER at such sites would warrant follow-up investigation. In contrast, ER within the fair range of ER or GPP rates would be less concerning at sites with low slope or low channel shading, because such sites have naturally higher rates of production and respiration. Over time UDWQ will continue to obtain metabolism from reference sites, which will allow us to remove the effects of important covariates in estimates of expected rates of GPP and ER. For instance, once sufficient reference site data have been collected, we will refine our assessment methods by repeating Random Forest models to predict background GPP and ER rates from natural environmental gradients, which will provide site-specific estimates of expected conditions and more refined metabolism assessments. In the interim, the fact that the upper end (fair-poor condition) thresholds are similar to those established from other independent studies provides confidence that these thresholds identify most sites with excess GPP or ER—particularly when coupled other nutrient response indicators.

Summary and Recommendations

In our study we quantified the relationship between nutrients, stream metabolism (GPP and ER). We identified thresholds for GPP and ER that can be used to quantify several condition classes based on these metabolism metrics. Perhaps most importantly, our comparisons of metabolism to DO criteria demonstrate that sites with excessive GPP and ER are also more likely to have issues with low DO. This correlation, coupled with the associated nutrient thresholds, will allow UDWQ to more accurately identify sites with potential DO problems. If DO impairments are identified, GPP and ER data will also provide insight into the nature of the impairment. For instance, if we observed excessively high ER at a stream with moderate or low GPP it would suggest that excessive carbon might be a more immediate nutrient-related concern than algal production. Alternatively, we might observe the converse—high GPP, but low ER—without a DO problem during summer months, which would prompt additional DO collections during periods of fall senescence.

The interpretation of slope and channel shading helps account for the influence of two important covariates, but these were regional analyses that can only identify patterns of how GPP and ER are broadly associated with nutrients. As we continue to apply metabolism data to other streams on a site-specific basis, UDWQ will consider and quantify the influence of covariates on stream metabolism processes. This is particularly important for site-specific investigations, which provide the best opportunity to explore local physicochemical controls on GPP and ER rates. Once these controls are understood, UDWQ can modify the regional thresholds to establish appropriate site-specific GPP or ER water quality goals. More important, insights gleaned from metabolism goals provide insight into remediation practices with the greatest potential to efficiently restore degraded ecosystems.

CHAPTER 5

ORGANIC MATTER STANDING STOCKS

Introduction

Biogeochemical links among nitrogen, phosphorus and carbon cycles in aquatic ecosystems have a rich history in ecological investigations (Redfield 1958). Anthropogenic increases in inorganic nutrients to streams have altered rates of accumulation, processing, storage and transport of organic matter at local to global scales (Webster et al. 1990, Webster and Meyers 1997). At large scales, alterations to organic matter dynamics affect atmospheric and oceanic carbon cycles (Kominoski and Rosemond 2012). At smaller scales, such as stream reaches, changes in storage and transport of organic matter impacts food resources (Hall et al. 2000, Hall and Meyer 1998), habitat availability (Walther and Whiles 2011, Yamamuro and Lambertii 2007) and ecosystem functions (Bilby and Likens 1980, Findlay et al. 2003).

Stream ecologists have focused on the role of organic matter budgets (Fisher and Likens 1973, Benstead et al. 2009), particularly those components that provide the energy base for stream food webs (Bonin et al. 2000, as reviewed in Tank et al. 2010). Early studies in organic matter budgets revealed that allocthonous organic matter (such as leaf litter) were the most important energy source in forested headwater streams (Fisher and Likens 1973). Whereas, other studies demonstrated that autochthonous energy sources become more important to food webs in larger, open canopy streams (Minshall 1978, Hall et al. 2000). Recent research has shown that organic matter derived from algae is more readily consumed by stream microbes than organic matter derived from terrestrial sources (Ylla et al. 2012, Lane et al. 2012), which highlights the importance of understanding specific carbon sources.

Dissolved nutrient concentrations are known to stimulate organic matter processing rates in streams (Triska and Sedell 1976, Robinson and Gessner 2000). The relatively high N and P content of heterotrophic bacteria and fungi compared to the N and P content of particulate organic matter suggests that increases in dissolved nutrients will stimulate microbial activity (Stelzer et al. 2003). In fact, a nutrient-mediated increase in microbial—primarily fungal—biomass has been observed following several experimental nutrient additions (Grattan and Suberkropp 2001, Gulis and Suberkropp 2002, Rosemond et al. 2002, Fereira et al. 2006, Sumberkropp et al. 2010) and respiration rates (Tank and Webster 1998, Young et al. 2008). These microbially mediated processes are a critical component of stream food webs because they convert more recalcitrant carbon, such as course particulate organic matter (CPOM), into more labile sources. Moreover, the microbes (i.e., bacteria and fungi) that colonize organic matter are a critical source of protein for macroinvertebrate shredders (Cummins 1974), and may be the principal way in which allochthonous carbon enters detrital food webs (France 2011).

Increases in heterotrophic productivity can result from several anthropogenic drivers including the quantity and quality external (allochthonous) organic matter inputs to the system (Stelzer at al. 2003), primary production and associated autochtonous organic matter stream inputs, or increases in the rate of organic matter processing resulting from inorganic nutrient inputs (Tank et al. 2010). Regardless of the source, increases in heterotrophic productivity have implications for stream oxygen dynamics. Oxygen dynamics in streams are controlled by physical and biogeochemical processes. Daily changes in stream dissolved oxygen (DO) concentration primarily result from the biological processes of gross primary production (GPP) and ecosystem respiration (ER) , and the physical process of reaeration (the exchange of gas between the stream and atmosphere):

$$\Delta DO = GPP - CR \pm reaeration$$

While decomposition of organic matter is a normal process in healthy streams, excess organic matter sometimes contributes to nighttime hypoxia, and less commonly anoxia, with deleterious effects to stream biota (Kemp and Dodds 2001, Connelly et al. 2004). The processes of oxygen consumption are often measured by the amount of oxygen consumed from the water column (biological oxygen demand (BOD)) or from the benthos (sediment oxygen demand (SOD)), or with reach-scale measures of stream metabolism (Acuna et al. 2004, Mulholland et al. 2001; see Chapter 4).

Rates of heterotrophic oxygen consumption are determined by the availability of nutrients (N and P), and on the availability and accessibility of organic matter (i.e. carbon). Therefore, organic matter standing stocks can be thought of as the potential for high rates of ecosystem respiration. Low minimum DO concentrations are offset by temporal site-specific factors (i.e. high reaeration related to flow or seasonally high GPP), which complicates the inference of potential problems with low DO from organic matter standing stocks. Nevertheless, the risk of low DO problems certainly increases with increasing organic matter. In addition, exceptionally high organic matter alters stream food webs, which can potentially degrade biological uses. As a result, measures of organic matter standing stocks have promise as functional indicators of stream condition.

In this study we investigated the relationship between water column nutrient concentrations (nitrogen and phosphorus) and organic matter standing stocks in stream ecosystems to determine if organic matter standing stocks increase in response to nutrient enrichment. Secondly, we investigated the relationship between organic matter standing stocks and in-stream oxygen dynamics including minimum dissolved oxygen concentrations and ecosystem respiration. We attempted to derive significant thresholds for each of the above relationships to develop indicators of the effects on nutrient enrichment on organic matter standing stocks that contribute to DO criterion violations.

Methods

Field Methods

We surveyed organic matter standing stocks at each of the 35 stream reaches (Figure 2.1 and Appendix A) between September and November in 2010. At each site we collected quantitative organic matter standing stocks from minimum 50m stream reach using a 100-count, point-intersect method (after Bowden et al. 2006; see Appendix X for detailed SOPs). We walked upstream at a 45 degree angle and sampled at five evenly spaced locations. At each point we recorded the channel unit type (rapid, riffle, glide or pool) and the substrate type (Fine Benthic Organic Matter (FBOM), gravel/sand, cobble, wood, macrophyte or filamentous algae). Two to seven replicate (reps) samples were collected for each channel unit-substrate combination sampled depending on relative abundance in the reach as follows: <10% = 2 reps, 10-39% = 3 reps, 40-59% = 4 reps, 60-79% = 5 reps, and 80-100% = 7 reps. Organic matter samples were collected using a stovepipe corer (FBOM, macrophytes and filamentous algae), syringe (sand/gravel) or scraping biofilm from a known area (cobble and wood). CPOM isolated from other stores with a 1 mm sieve. Organic matter samples were placed in Whirl-pak® bags, placed on ice, and then filtered (FBOM, gravel/sand, cobble and wood only) on pre-ashed Whatman GF/F filters and frozen within 16 hours of collection.

Surface water nutrients were collected on a minimum of three distinct sampling dates at the upstream and downstream end of the reach between July and October and were analyzed at the Aquatic Biogeochemistry Laboratory at Utah State University (Valderrama 1981). Samples across time and location (minimum six) were averaged to determine water column nutrient concentrations for the reach. At each site we also deployed a water quality probe (YSI 6600V2 or 600 OMS V2) to measure DO at five minute intervals for a minimum of 48 hours at the downstream end of each reach.

Laboratory Methods

Organic matter samples were subsequently quantified as ash free dry mass (AFDM) following established laboratory methods (Kiry et al. 1999). Samples were dried at 60° C for 48 hours before being weighed on an analytical balance (Denver M-220, to 0.0001g). Samples were then combusted at 450° C for 2 hours, re-wetted, dried and reweighed. AFDM was computed as the difference between the mass prior to ashing (organic plus inorganic matter) and subsequent to ashing (just inorganic matter). Larger samples of macrophytes and filamentous algae were subsampled prior to combustion and then scaled up to the entire sample. Multiple samples for each channel unit – substrate type were averaged. Organic matter areal concentrations were multiplied by the relative abundance of each channel unit–substrate type combination at the stream reach and then recorded as relative abundance per m^2.

Analytical Methods

POOLING REACH-SCALE DATA FOR REGIONAL ANALYSES

Our collection methods for OM standing stocks were fairly detailed—differentiating among several stream habitats and OM pools. These methods provides fairly accurate reach-scale estimates of OM standing stocks because they intrinsically incorporate among stream differences in OM sinks — (they are scaled from the spatial extent of pools and riffles) and sources (different standing stocks are collected separately within each habitat). Such procedures will ultimately provide insight into the relative importance of OM sources and sinks at local scales. However, these analyses primarily aim to establish more general regional relationships between nutrients and OM standing stocks. As a result we opted to combine the laboratory results of reach-scale OM stores in a couple of ways. First, we focused on OM stores that are most closely coupled to autchthonous production, including: filamentous algae, and the biofilm samples obtained from wood (epixylon), rock/hard surfaces (epilithon), and sand (episammon). We also included Fine Benthic Organic Matter (FBOM), which ultimately could have come from either authochthonous sources or represent processed OM from allochthonous sources, because these stores are important determinant of microbial and fungal heterotrophic production which potentially can be directly influenced by inorganic nutrients.

Equally important, were decisions about what OM standing stocks to exclude: we did not include contributions from macrophytes or coarse benthic organic matter (CBOM) in our total organic matter estimates. The presence and abundance of stream macrophytes is strongly influenced by physical factors such as flow regimes or substrate size (Chambers et al. 1991). In fact, macrophytes were entirely absent or in very low abundance at ~90% of these streams. Three of the four exceptions were located within the East Canyon Creek drainage, where the WWTP already limits nutrients to meet TMDL expectations. More broadly, some marcophytes also have the ability to assimilate nutrients from the water column and from sediments and the relative contribution of each can change depending on site-specific nutrient availability and macrophyte species (Madsen and Cedergreen 2002, Barko and Smart 1981). As a result of these factors, macrophyte abundance at sites in this study was not significantly associated with nutrients (Table 5.2). Therefore, including

Table 5.1 Relative amount of organic matter standing stock stores for each site, expressed relative to all organic matter at the site (%) and by mass (AFDM). Site names and descriptions can be found in Table 2.1.

	Reach Wide	FBOM		Epixylon		Epilithon		Episammon		Fil. Agae			
Site	$gAFDM/m^2$	$gAFDM/m^2$	%	$gAFDM/m^2$	%	$gAFDM/m^2$	%	$gAFDM/m^2$	%	$gAFDM/m^2$	%	Mean TN	Mean TP
BEC-AB	41.0	34.1	83.1	3.7	9.0	0.0	0.0	0.0	0.0	3.3	8.0	0.404	0.058
BEC-BL	256.0	252.1	98.5	3.2	1.3	0.0	0.0	0.0	0.0	0.7	0.3	6.351	0.838
BLACKFK	15.3	1.9	12.4	0.0	0.0	7.3	47.6	4.1	26.7	2.0	13.0	0.188	0.008
DCSP-AB	132.0	128.7	97.5	1.9	1.4	0.0	0.0	0.0	0.0	1.3	1.0	2.216	0.135
DCSP-BL	118.4	118.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.286	1.894
DIAFK	17.9	1.5	8.6	0.0	0.0	9.3	52.1	7.0	39.2	0.0	0.0	0.410	0.084
FISHCK	38.8	20.0	51.6	0.0	0.0	9.4	24.2	8.0	20.6	1.4	3.6	0.246	0.063
HUNTCK	23.8	1.2	5.0	0.0	0.0	22.2	93.4	0.0	0.0	0.4	1.7	0.455	0.076
KIMBALL	68.5	63.1	92.1	0.0	0.0	5.3	7.7	0.2	0.3	0.0	0.0	0.279	0.028
LBRAVON	27.0	5.1	18.7	0.0	0.0	19.0	70.3	2.9	10.7	0.0	0.0	0.343	0.021
LBRW-AB	16.0	12.5	78.2	0.0	0.0	2.9	18.2	0.6	3.8	0.0	0.0	1.175	0.075
LBRW-BL	7.8	3.8	48.7	1.1	14.1	1.8	23.1	1.0	12.8	0.1	1.3	1.085	0.084
LOGR1000	22.8	4.2	18.3	0.0	0.0	1.2	5.3	17.3	75.8	0.1	0.4	0.139	0.012
LOGRDUG	10.3	0.3	2.5	0.0	0.0	8.8	85.7	1.3	12.7	0.0	0.0	0.424	0.023
LOGRTB	5.4	1.6	29.6	0.0	0.0	1.7	31.6	0.2	3.7	1.9	35.3	0.132	0.012
MRTRE-AB	20.4	15.6	76.5	1.7	8.3	2.0	9.8	1.1	5.4	0.0	0.0	2.828	0.236
MRTRE-BL	44.9	40.2	89.5	3.8	8.5	0.5	1.1	0.5	1.1	0.0	0.0	3.897	0.445
NFCHLK	8.8	0.2	1.9	0.0	0.0	8.4	95.1	0.2	2.3	0.0	0.0	0.161	0.006
PRICER	13.8	4.5	32.7	0.0	0.0	0.0	0.0	9.0	65.2	0.3	2.2	0.388	0.050
PRP-AB	19.7	13.2	67.1	6.4	32.4	0.0	0.0	0.1	0.5	0.0	0.0	0.710	0.289
PRP-BL	38.0	37.1	97.8	0.2	0.5	0.5	1.3	0.1	0.3	0.0	0.0	2.950	0.732
SALTCK	5.6	1.0	18.0	0.0	0.0	3.7	66.4	0.2	3.6	0.7	12.6	0.154	0.016
SCSNYD-AB	4.2	4.0	94.6	0.0	0.0	0.0	0.0	0.2	4.7	0.0	0.0	0.319	0.015
SCSNYD-BL	52.6	1.9	3.7	0.0	0.0	0.0	0.0	0.4	0.8	50.3	95.6	14.717	2.212
SFKLBR	13.5	10.8	79.4	0.0	0.0	1.9	14.0	0.9	6.6	0.0	0.0	0.229	0.017
SPRFV-AB	28.1	14.8	52.7	0.0	0.0	6.1	21.7	1.1	3.9	6.1	21.7	1.324	0.019
SPRFV-BL	30.7	11.1	36.3	0.0	0.0	10.2	33.2	8.6	28.0	0.8	2.6	1.677	0.078
SPRM-AB	60.4	13.8	22.7	0.0	0.0	1.1	1.8	10.0	16.5	35.6	58.9	1.232	0.033
SPRM-BL	194.5	165.7	85.2	0.0	0.0	0.0	0.0	28.7	14.8	0.0	0.0	10.416	7.897
TIEFK	37.5	17.1	45.5	0.0	0.0	0.1	0.3	18.6	49.6	1.8	4.8	0.118	0.007
UKMURD	34.1	5.2	15.2	0.0	0.0	23.6	69.2	5.0	14.7	0.3	0.9	0.109	0.004
UPRNFK	9.0	0.0	0.0	0.0	0.0	8.9	98.5	0.1	1.1	0.0	0.0	0.113	0.003
WEBR	43.6	0.6	1.3	0.0	0.0	41.5	95.1	1.5	3.4	0.1	0.2	0.382	0.040
WROAK-AB	19.0	1.4	7.3	0.0	0.0	12.0	63.1	3.8	20.0	1.8	9.5	0.109	0.009
WROAK-BL	15.0	0.4	2.5	0.0	0.0	5.7	38.0	1.7	11.3	7.2	48.0	0.125	0.020

macrophytes in these regional analyses would add unnecessary noise to the organic matter-water column nutrient relationships. We also excluded CBOM because it is made up of mostly organic matter terrestrial in origin (small branches, leaves, etc.), which at least until processed are more important source of energy for invertebrate that heterotrophic microbes and fungi (Wallace et al. 1982, Cuffney et al. 2006). Moreover, the relative importance of invertebrates in organic matter processing decreases from upstream to downstreams, which confounds the ability to measure relationships with nutrients on OM standing stocks. Finally, these investigations were conducted in the summer, which misses autumn inputs, which is the most important period of heterotrophic inputs among temperate streams. In fact, CPOM organic matter comprised less than 6% of the OM standing stocks in late summer at our study streams. As a result, there were no significant relationships between nutrients and CPOM among these study streams (Table 5.2). As with macrophytes, the decision to exclude these pools from regional analyses does not suggest that these sources of OM are unimportant, they are; however, in this case we did not have sufficient information to evaluate they components for purposes of our stress-response analyses.

We pooled the remaining organic matter stores of interest based on their relative abundance in reach stream to obtain reach-scale abundance estimates. First we calculated the relative abundance of each organic matter pool within riffle and then pools within each reach. Next we scaled these data according to the spatial extent of these habitats to generate reach-scale abundance estimates. Finally, we compiled these reach-scale estimates to obtain the abundance of OM stores of interest for each stream.

RELATING STREAM NUTRIENTS TO ORGANIC MATTER

For these key OM stores, we evaluated the extent to which OM standing stocks were correlated with either TN or TP concentrations to help directly link the OM investigations to other indicators. We used simple linear regression on log(x) transformed data to identify any linear relationships between nutrients and organic matter standing stocks. Next, we used nonparametric deviance reduction to identify TP and TN thresholds that separated organic matter standing stocks into distinct groups with maximal within group similarity in the mean and variance of chemistry samples (Qian et al. 2003, package `rpart`). Significance of multiple threshold models was tested using ANOVAs followed by a Tukey's Honestly Significant Difference (HSD) test.

RELATING OXYGEN DYNAMICS TO ORGANIC MATTER

Table 5.2. Minimum and 30-day average dissolved oxygen standards (mg/L) for three designated aquatic life beneficial use (UAC R317-2-14): coldwater fish (3a), warmwater fish (3b) and nongame fish (3D).

Aquatic Life	Minimum DO Standard	30-day Average DO Standard
Coldwater Fish	4.0	6.5
Warmwater Fish	3.0	5.5
Nongame Fish	3.0	5.0

We also evaluated the relationship between organic matter standing stocks and in stream oxygen dynamics. If increased organic matter standing stocks are consistently associated with violations of DO then we have a potential causal link to degradation of aquatic life uses. We used DO measurements taken every five minutes over a 48-72 hour period to calculate the observed minimum DO (mg/L). With two exceptions we used the aquatic life uses assigned to each site to evaluate percent of samples that exceeded Utah UDWQ's minimum and 30-day average DO criteria (UAC R317-2-14, Table 1). Two sites were designated with habitat-limited uses (3D), which do not have DO criteria, and DO criteria for 3C uses were used for these sites.

RELATING ORGANIC MATTER TO ECOSYSTEM RESPIRATION

We also compared OM standing stocks with Ecosystem Respiration (ER, see Chapter 4), because excess carbon is potentially as important as TN and TP in causing low DO within streams. As with nutrients, we used nonparametric deviance reduction techniques to determine thresholds of organic matter standing stocks that best identify sites that violate DO criteria and ER values of potential concern. The significance of thresholds identified via these deviance reduction analyses were then tested with student's t-tests or Wilcoxon Rank Sum tests if statistical assumptions of parametric methods were violated.

EVLAUATING THE INFLUENCE OF PHYSICAL COVARIATES

For the purpose of these analyses our organic matter sampling methods attempted to focus of the storage of autochthonous organic matter (i.e. excluding CBOM), because these are the carbon stores that are directly influenced by nutrient inputs. Even so, a large portion of the stored organic matter within streams may be derived from terrestrial sources and stored as FBOM or incorporated into biota. As a result, organic matter standing stocks will vary, not only with nutrients, but also with a number of physical factors unique to the watershed and riparian corridor. Hence, we evaluated the relative importance of nutrient concentrations and watershed characteristics on among-stream differences in organic matter standing stocks.

We fitted Random Forest regression models (Breiman 2001, package `randomForest`) to predict organic matter standing stocks from stream nutrients (TN and TP) and other chemical constituents, watershed characteristics derived from USGS Stream Stats (USGS 2014), and site-specific physical habitat measures from UDWQ's UCASE protocols (Table 5.4, UDWQ 2012). We then compared the relative signal from nutrients and the most important physical covariates (determined from Random Forest regression) to separate nutrient-organic matter standing stocks relationships from other potential sources of variation.

Random Forest regressions were used to identify the suite of candidate covariables that best predict among stream differences in organic matter standing stocks. Random Forest regressions builds multiple successive regression trees using "bagging", which roughly equates to more common bootstrapping procedures. Specifically, multiple regression trees are constructed, each one based on a subsample of observations and a subset of available data (Breiman 1996). In each case, regression trees are built with the best possible splits among the subset of predictor variables to obtain the most accurate regression tree possible. Each tree then gets a vote in the final prediction, which is based on the ensemble of trees—the forest. Once forests were constructed, we determined the relative importance of potential predictor variables with the variable importance (VarImp) procedure within the Random Forest package. The VarImp procedure randomly reassigns (shuffles) observations for each potential predictor one at a time before rerunning the forest models. The increase in error (MSE) that results from the new RF regression quantifies the relative importance of that variable to the Random Forest regression. This procedure is then repeated for each candidate variable, one at a time. Once complete the relative importance of predictors can be determined. Those variables that lead to the greatest reduction in predictive accuracy once observations have been randomly assigned are assumed to be of greatest importance. We choose Random Forest models to determine variable importance because these approaches are unbiased by datasets of highly correlated variables, robust against over fitting (due to the randomization procedures), and do not require adherence to parametric statistical assumptions (i.e., homoscedasticity, normal distributions) (Breiman 2001).

It is important to explore potential bias among all covariates, yet we also wanted a model that was readily interpretable and parsimonious. As a result, a second Random Forest model was fitted with only the most important variables. We reasoned that if a second model with fewer variables performed as well as the larger model, then it would be easier and more cost-effective to measure important covariates in future management applications. All analyses were conducted in R v2.15.0 (R Core Development Team, 2012).

Table 5.3. Spearman rank correlation coefficients between total nitrogen and total phosphorus and each of the organic matter storage compartments evaluated in this study.

		Spearman Rank Correlation Matrix						
		FBOM	CBOM	Epixylon	Epilithon	Episammon	Macrophytes	Fil Algae
TN	r =	0.56	0.06	0.45	-0.47	-0.24	0.14	-0.14
	p =	<0.001	0.730	0.007	0.005	0.157	0.409	0.417
TP	r =	0.56	0.23	0.52	-0.45	-0.24	0.05	-0.17
	p =	0.005	0.185	0.001	0.006	0.173	0.786	0.335

Results

General Patterns with Organic Matter Standing Stocks

Organic matter standing stocks varied by thee orders of magnitude among all sites with total stores ranging from 4.24 to 256.0 g AFDM/m^{2,} and a median of 23.76 g AFDM/m^2. Fine benthic organic matter (FBOM) was the largest contributor to reach wide organic matter with an average of 45% of the total standing stock followed by epilithon (30.5%), episammon (13.1%), filamentous algae (9.2%) and epixylon (2.2%)(Table 5.1).

Relationship between Organic Matter and Nutrients

Among-steam nutrient concentrations were also quite variable with a log-normal distribution. Nitrogen (TN) varied from 0.109 to 14.72 mg/L, with a median of 0.404 mg/L. Phosphorous (TP) varied from 0.003 to 7.89 mg/L, with a median value of 0.04 mg/L. After log transformation, we found significant relationships between organic matter standing stocks and TN using linear regression (r^2 = 0.40, p <0.001) and TP (r^2=0.39, p <0.001, Figure 5.1).

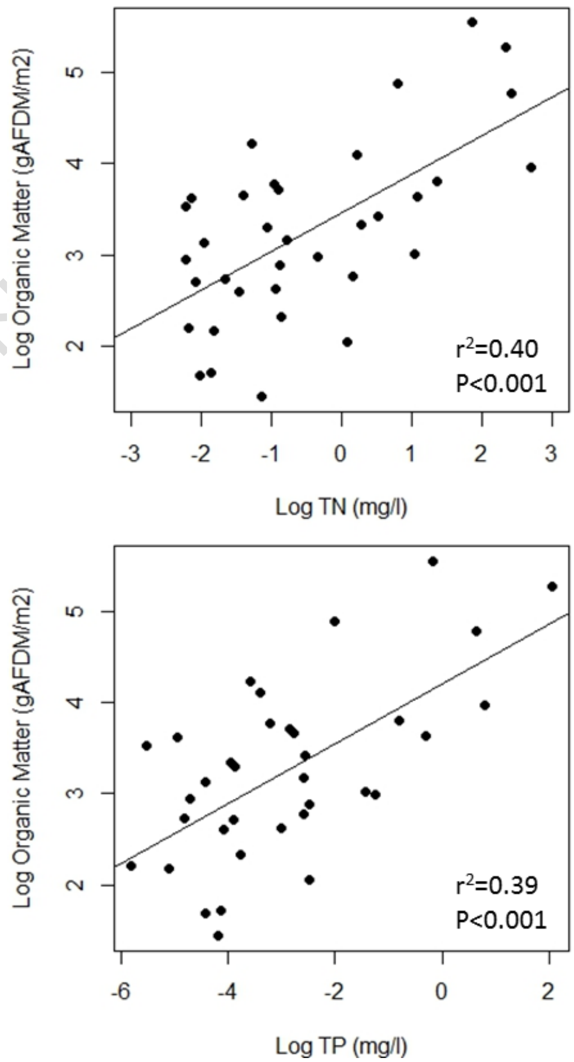


Figure 5.1. Linear regression between surface water total nitrogen (top panel, r^2=0.40, p<0.001) and total phosphorus (bottom panel, r^2=0.39, p<0.001) and AFDM (g/m^2).

We also evaluated the extent to which different stores of organic matter were related to nutrients and found that among all sites only FBOM, epixylon, and epilithon varied significantly with TN and TP (Table 5.2).

Identification of Nutrient Thresholds

We used nonparametric deviance reduction to identify thresholds of TN and TP that best group organic matter into distinct groups. For TN we found thresholds at 0.238 and 1.95 mg/L TN that separated organic matter standing stocks into three groups (L-low, M-medium and H-high) with statistically different nitrogen concentrations (ANOVA p=<0.001, Tukey HSD L-M p=0.26, L-H p<0.001, M-H p=0.002, Figure 5.2). Using the same procedure for TP we found statistically significant thresholds at 0.026 and 0.589 mg/L (ANOVA p<0.001, Tukey HSD L-M p=0.01, L-H p=<0.001, M-H p=0.006, Figure 5.2).

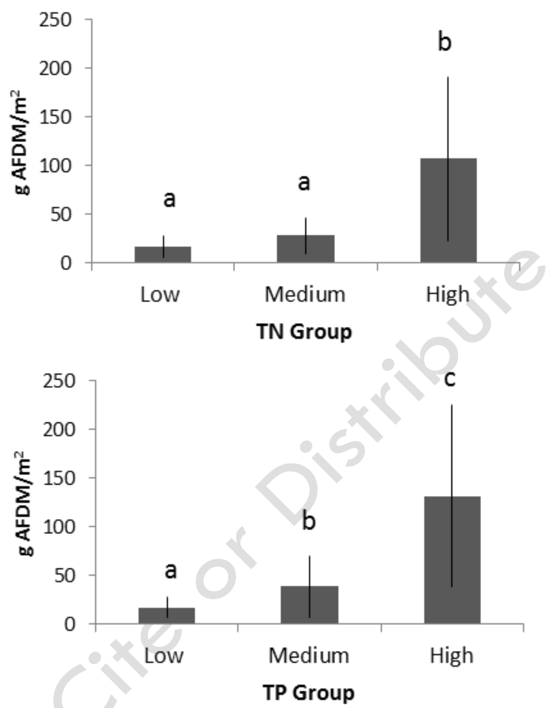


Fig 5.2. Organic matter (AFDM (g/m^2)) standing stocks among streams within low, medium and high nutrient groups. Letters indicate significance (ANOVA and *post-hoc* Tukey HSD) and error bars are one standard deviation.

Relationships with Existing Dissolved Oxygen Criteria

OM standing stocks were strongly associated with DO concentrations obtained from 48-72 hours probe deployments. Minimum DO values among all sites varied from 0.39-8.53 mg/L. Overall 31 of the 35 sites never exceeded the minimum (acute) DO criterion, but in the 4 sites where exceedences did occur DO values were below the criterion for ~1/3 of the day. DO fell below the 30-day average (chronic) DO criterion at 14 of the 35 stream sites. Among the 14 streams where DO fell below chronic thresholds the amount of time the streams remained below the criterion varied considerably, from 3% to 75% of 5-minute interval samples. Daily Ecosystem respiration (ER) rates were calculated for 31 sites (see Chapter 4 for details) and were significantly, albeit weakly, correlated to OM standing stocks (linear regression r^2 = 0.15, p =0.02, data not shown).

We were also interested in determining whether sites above and below DO benchmarks could be distinguished with organic matter standing stocks. To do this we used NDR to identify independent thresholds of organic matter standing stocks for each of the four DO benchmarks. In all four cases, we found a threshold of 48.76 g AFDM/m^2. Each of these thresholds divided streams into groups with statistically significant ($p<0.05$) differences in OM standing stocks (Figure 5.3)

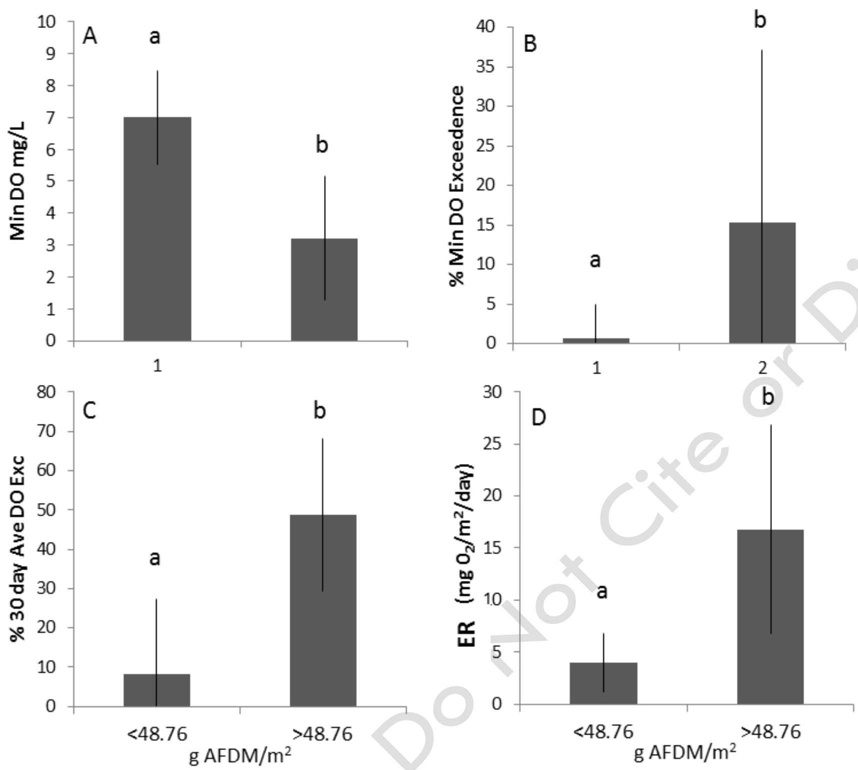


Fig 5.3. Relationship among several water quality benchmarks and among streams within with low (<48.76 g AFDM/m^2) and high Organic matter. Data for the water quality relationships were obtained from a minimum 48 hour sample period. Panels A depicts the minimum DO observation of this period. Panels B and C depict the number of five minute observations that fall below two numeric DO criteria that differ with respect to their averaging periods. Panel D depicts compares ecosystem metabolism among streams with low and high organic matter. Letters indicate significant differences determined by student's t-tests (A and D) or Wilcoxon rank sum test (B and C). Error bars are one standard deviation.

The Influence of Physical Covariates

The Random Forest model based on all candidate variables was significant, albeit with fairly low accuracy (pseudo $r^2=0.239$). Nevertheless, the model did elucidate the relative importance of the variables that we evaluated. TN and TP scored high on measures of variable importance (68.4% and 53.8% increase MSE from VarImp procedure). The watershed characteristic percent fast water habitat (% riffles + rapids) was the most important variable with an 87.4% increase in MSE. Two reach-scale physical habitat parameters, mean wetted width (m) and watershed area (miles^2), were about half as important as nutrients (36.1% and 26.0% increase MSE, respectively) in explaining

Table 5.4 Physical, chemical and landscape-level descriptor of environmental gradients that were evaluated to estimate the relative importance of nutrients and other stream characteristics. The importance of variables was determined from the % increase Mean Squared Error (MSE) outputs of Random Forest models. Higher MSE indicates that when values in a variable were randomized the model performance declined. Data were obtained from the Utah State University Aquatic Biogeochemistry Laboratory (USU ABL), Utah Unified Public Health laboratories (UPHL), U.S. Geological Survey Stream Stats program (USGS) or the Utah Division of Water Quality Comprehensive Assessment of Stream Ecosystems program (UDWQ).

	Units	Full Model %Increase MSE	Best Subset Model %Increase MSE	Source
Total Nitrogen	mg/l	68.4	89.1	USU ABL
Total Phosphorus	mg/l	53.8	81.7	USU ABL
Turbidity	NTU	15.5		UPHL
Total Suspended Solids	mg/l	20.6		UPHL
Slope	%	19.9		USGS
Basin Area	mi^2	26.0	39.4	USGS
Herbaceous Upland	%	5.8		USGS
Forested Watershed	%	2.2		USGS
Basin Slope	%	0		USGS
Mean Water Depth	cm	12.4		UDWQ
Mean Wetted Width	m	36.1	14.4	UDWQ
Mean Thalweg Depth	cm	13.6		UDWQ
Bankfull Height	cm	-4.8		UDWQ
Channel Incised Height	cm	1.8		UDWQ
Channel Width:Depth Ratio		14.5		UDWQ
Fast Water Habitat	%	87.4	137.9	UDWQ
Riparian Corridor Bare Ground	%	-8.5		UDWQ

among stream OM differences. We fitted a second Random Forest model with only these five variables and model performance actually increased (pseudo r^2=0.431). The final model ranked variable importance in identical order as the first model with the full suite of parameters. Percent fast channel (123.9% MSE) was most important followed by TN (97.0% MSE), TP (79.2% MSE), watershed area (42.9% MSE) and mean wetted width (24.0% MSE) (Table 5.4).

Discussion

Organic Matter Standing Stocks

We found a significant relationship among organic matter standing stocks and nutrient concentrations (TN and TP) with relatively strong linear relationships with both TN ($r^2 = 0.40$) and TP ($r^2 = 0.39$), especially considering the diversity of streams sampled. This confirms that the sampling methods chosen specifically to detect the portion of organic matter stocks that are influenced by in-stream nutrients (i.e. excluding CBOM) was successful. Of all of the storage compartments evaluated for organic matter standing stocks FBOM was the largest. While FBOM is certainly an important sink of organic matter in streams it may also produce some of the error in the relationships with in-stream nutrients or the ER responses. Given that these data were collected in autumn, some of the FBOM was undoubtedly from autochthonous sources. However, the FBOM undoubtedly also contained carbon of terrestrial origin that was broken down into smaller particles and stored as FBOM. We would not expect terrestrially-derived FBOM to be associated with stream nutrients except in the sense that high nutrient streams may have lower allochthonous:autochthonous FBOM due to increases in carbon processing rates. The fact that ER was higher in high carbon streams suggests that may be occurring within these systems. In this context, the relationships between nutrients and OM would nutrients would be in the opposite direction of autochthonous sources, which would decrease the strength of these relationships. Increases in carbon processing rates have been directly quantified by others. For instance, Benstead et al. (2009) showed that artificial N and P additions to detritus fed streams increased processing rates of CBOM to FBOM. This indicates that although we can't be sure of the origin of FBOM (terrestrial or aquatic) in our study sites, the high levels of FBOM may be indicative of responses to nutrient enrichment. Future OM research by UDWQ will quantify the $^{13}C/^{12}C$ of FBOM, which may help better elucidate the ultimate source of carbon inputs (Palmer et al. 2001). Nitrogen isotopes may also be useful because aquatic plants and algae are typically enriched by 3% $^{15}N/^{14}N$compared to terrestrial counterparts (French 1995).

Nutrient Thresholds

We determined thresholds of TN and TP that best separated organic matter standing stocks into three groups of streams. On average, a TN concentration of 0.238 mg/L distinguished between streams with low to moderate levels of OM, whereas a concentration of 1.95 mg/L distinguished between streams with moderate and high levels of OM. For TP the lower threshold was 0.026 mg/L, where 0.589 mg/L defined the upper threshold. We found that differences between organic matter standing stocks were less distinct between the Low and Medium nutrient groups than for the Medium and High nutrient groups. In fact, in the case of nitrogen, organic matter standing stocks did not differ between TN-Low and TN-Medium groups (Figure 2). UDWQ will use these thresholds, in conjunction with others, to identify sites with potential nutrient-related problems. However, these nutrient thresholds are among the weakest of responses that we investigated perhaps because we were unable to determine whether FBOM was of allochthonous and autochthonous origin.

OM and DO

If UDWQ intends to use OM standing stocks as an indicator of anthropogenic eutrophication, it is important that any thresholds identified are *both* statistically significant *and* representative of potentially deleterious effects on aquatic life uses. To examine the ecological relevance of our thresholds, we compared organic matter stocks to DO metrics (minimum DO and ecosystem respiration) and Utah's DO water quality criteria (minimum DO standard and 30-day average DO standard).

The majority of sites (89%) did not show a violation of the minimum DO criterion during the study period. Among sites where DO fell below water quality benchmarks, period of low DO were long-lived, which makes it highly likely that DO threatens aquatic life at these locales. Importantly, UDWQ likely would have missed even these extreme circumstances with routine grab samples, because standards were not exceeded for 2/3 of the day at times where grab samples are most likely to occur. Ideally, DO assessment decisions would always be made with quality, high frequency data, yet these data are not always readily available for assessment purposes. Another important consideration when interpreting these DO results is that they only capture a 3-7 day snapshot of DO conditions. Periods of low DO may occur during other times of the year, particularly at sites where clearly living OM standing stocks are high because ER would increase during autumn senescence as living OM becomes more labile; moreover, the ability of GPP to offset losses of DO from ER may also be diminished during these period due to decreases in temperature and algae abundance A more temporally stable surrogate measure, like OM standing stocks, may help identify sites with potential DO problems.

Aquatic biota can also be negatively affected by chronic—long-term—exposure to low DO. We estimated chronic effects of low DO by comparing the percent of times DO fell below the 30-day average minimum DO criterion and subsequently relating these data to OM standing stocks (Table 5.1). Samples at sites with organic matter >48.76 g AFDM/m^2 were lower than their 30-day DO criterion 48.8% of the time, whereas this only occurred 8.8% of the time at sites with low organic matter standing stocks. These relationships do not necessarily imply impaired conditions because our data are temporally limited. Some sites that are fully attaining aquatic life uses may occasionally fall below the 30-day average criterion without harm to aquatic life uses. Nevertheless, these data provide credence to the use of OM as a potential screening tool for sites with potential DO problems, especially considering that OM thresholds were identical for other DO water quality benchmarks that we evaluated.

OM and Metabolism

As expected, we found that sites that had greater organic matter standing stocks had higher rates of ecosystem respiration. In fact, ER organic matter thresholds were identical to those derived from DO water quality criteria. While we did not expect identical organic matter thresholds among these indicators, it is not surprising that they are related. All four of our oxygen metrics were not completely independent as they were calculated from the same dissolved oxygen record. Eutrophication, by definition, increases autotrophic production and biomass in aquatic systems. A less frequent, but potentially equally important consideration is that nutrient enrichment can also stimulate heterotrophic productivity and increases in organic matter processing rates (Robinson and Gessner 2000), which further reduces DO in streams. In fact, in the Benstead et al. (2009) study mentioned above, they found increased heterotrophic respiration rates per gram of substrate on leaf litter, woody debris and FBOM after two years of whole stream nutrient additions. This study agrees with previous literature that has shown a positive effect of nutrient enrichment on microbial respiration (Gulis and Suberkropp 2002, Cross et al. 2006). Our study suggests that heterotrophic responses to nutrients may hold true across a broad range of streams with different organic matter composition and stocks.

Physical Covariates

Numerous stream physical characteristics affect the delivery, storage, and transport of OM within streams. The most import variables identified with Random Forest regression models are consistent with OM processes. Among all of the variables evaluated, we found that percent % fastwater habitat (transport), nutrients (TN, TP, storage/processing), and stream size (watershed area and mean wetted width, delivery and storage) were the strongest predictors of organic matter storage. These observations are consistent with accepted stream ecology theories such as the River Continuum Concept (RCC, Vannote et al. 1980) that predicts changes in physical and ecological characteristics from headwaters to the valleys. Increases in water velocity (likely due to increased slope) are all likely to occur at higher order streams where the majority of organic matter inputs are from finely processed terrestrial matter (CBOM). All of the important physical covariates quantify changes from smaller headwater streams to the valleys. The influence of covariates does not preclude the potential importance of TN and TP. The Random Forest models suggest that both of these macronutrients have a more important influence on organic matter standing stocks than stream size, which in turn suggests that biogeochemical processes, not just stream size, are important factors influencing the storage and retention of OM within streams.

Summary and Recommendations

We established relationships between nutrients and OM and between OM and DO dynamics. Significant thresholds were developed for each of these relationships that can help guide future monitoring and assessment efforts, ultimately helping to identify potential solutions to streams with nutrient-related impairments. We demonstrated that increased OM standing stocks are associated with lower minimum DO concentrations, and also more violations of DO criteria, which provides a link between OM indicators and support of aquatic life uses. Our multivariate analyses show that among stream differences in OM standing stocks could not be explained by physical characteristics alone, and that excess nutrients play a role in organic matter accumulation and storage. We suggest a response OM standing stock threshold of >48.76 g AFDM/m^2 as a broadly applicable regional indicator of nutrient enrichment.

In reality, follow-up organic matter studies will be most useful to inform site-specific investigations among stream where low DO concerns have been identified because high frequency DO data are increasing inexpensive and easy to collect. In cases where DO concerns have been identified the OM thresholds could be used to infer whether or not organic matter pools are atypically high. Insights gleaned from these investigation will prove insightful, for instance, in TMDL studies addressing low DO problems. In this circumstance, OM standing stocks may be a useful water quality objective—provided that local conditions such as slope are accounted for—because it provides a time integrative measure of an important driver of low DO. Also, in some circumstances reductions in OM standing stocks may be a direct measure of BMPs that may be employed to address DO concerns, which may make this indicator a robust indicator of incremental progress.

Of course, these data also highlight the importance of confirming these indicators on a site-specific basis, which will remain an integral part of Utah's nutrient reduction efforts. Additional efforts will need to further elucidate on the relative role of nutrients and important covariates. Specifically, follow-up investigations will need to address the site-specific importance of habitat characteristics associated with organic matter transport (i.e., slope and habitat complexity). It will also be important that these site-specific investigations also identify important organic matter sources. Additional details provided by these site-specific investigations will more accurately characterize the causes of DO impairments and the influence of increased nutrients on stream food webs. Such insights, together with other lines of evidence discussed in this report, should future remediation efforts.

Relationships among Nutrients, Nutrient Responses, Aquatic Life and Recreation Uses

SECTION 2

Chapter 6

STRUCTURAL INDICATORS: RELATIONSHIPS BETWEEN NUTRIENTS AND STREAM BIOTA

Introduction

Worldwide, aquatic resource managers use numerous measures of stream condition to identify water quality problems. Bioassessments—quantitative descriptions of anthropogenic alterations to the composition or structure of aquatic assemblages— are among the most meaningful assessment tools. Resident aquatic communities integrate the effects of stressors through time because they are subjected to long-term impacts to ecosystems (weeks to years). Biological assessments also integrate the effects of multiple stressors, both spatially and temporally (Fausch et al. 1990). They are also of direct interest to the public, and this interest is expressed in the goals of many regulations that seek protection of aquatic resources. In the United States, the support and maintenance of biological integrity is one of the fundamental objectives of the Clean Water Act (CWA §101(a)).

Biological Integrity: An Objective of the Clean Water Act
The Clean Water Act seeks to protect, maintain and restore the biological integrity of the nation's waters. This objective is built upon the concept that increasing human activity alters stream function and structure. Currently the generally accepted definition of biological integrity (after Frey 1977), is "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region." No single measure can capture all of the ecological attributes captured in this definition, but various indicators have been established to estimate the degree of departure based on alterations to the composition of stream biota.

Excess nutrients are one of the greatest threats to the biological integrity of the nation's waters (USEPA 2000). Anthropogenic sources of nutrients to the Nation's waterways have been a known stressor to aquatic communities prior to the adoption of the Clean Water Act (e.g., Carr 1962, Vallentyne 1974). Aquatic biological communities, especially diatoms and macroinvertebrates, are particularly sensitive to excessive surface water nutrients (King and Richardson 2003, Smith et al. 2007, Wang et al. 2007, Van Sickle and Paulsen 2008). At very high concentrations, nutrients—particularly nitrate and ammonia— can be acutely toxic to biota, but such conditions rarely occur in streams. More often, human nutrient inputs to streams alter ecosystem processes (i.e., dissolved oxygen fluctuations, organic matter processing), which in turn affects the structure and condition of biotic food webs. Structural indicators of nutrient enrichment provide quantitative estimates of the extent of alteration to biological assemblages.

Not all species respond similarly to increased nutrients; some species cannot tolerate changes caused by excessive nutrients, whereas others are adapted to such conditions and actually thrive in nutrient enriched conditions (Davies and Jackson 2006). Hence, it is important that metrics used to identify structural responses to nutrient enrichment account for orthogonal responses among taxa. Moreover, structural responses vary among sites with different environmental conditions (i.e., channel shading, temperature, and substrate characteristics), and bioassessments need to account for the influence of these covariates in order to accurately estimate the composition expected under unaltered conditions.

Different assemblages (i.e., diatoms vs. macroinvertebrates) can also differ with regard to their relative sensitivity to nutrient enrichment. Some assemblages, like algae and diatoms, respond directly to nutrients, whereas responses for others are indirect. For instance, as nutrients at a site increase diatoms are often replaced by other algae taxa, which can then affect invertebrates—either positively or negatively—due to the resulting changes in habitat or food quality (i.e. diatoms to filamentous algae) (Peterson et al. 1993). Compositional changes are not independent, so it is not necessarily appropriate to give deference to one assemblage over another. Instead, regional thresholds are best derived from the weight of evidence obtained from several different measures of condition. Ideally, they would reflect several aspects of food web dynamics (i.e., different trophic levels, functional feeding groups).

Measures of biological integrity intrinsically involve multivariate comparisons of compositional similarity. It is preferable to examine several analytical approaches when developing quantitative biological assessment tools. Convergence among thresholds obtained from different analytical techniques improves confidence that thresholds are environmentally meaningful and not simply statistical artifacts.

The primary objective of this chapter is to report on N and P concentrations found to best explain among-stream differences in the composition of diatom and macroinvertebrate assemblages. First, we evaluated whether different taxa (i.e., genera or species) responded systematically to varying nutrient concentrations using several analytical approaches. First, we analyzed individual diatom and macroinvertebrate taxa that consistently increase or decrease in abundance among streams with varying nutrient concentrations. To identify these taxa we used Threshold Indicator Taxa ANalysis (TITAN, Baker & King 2010, King and Baker 2010), a recently proposed analytical approach derived from long-established ecological relationships. TITAN integrates bidirectional taxa occurrences (presence vs. absence) and relative abundance relative to an environmental stressor to determine stressor-response thresholds. TITAN identifies thresholds for individual taxa, but also integrates these thresholds into single thresholds to establish a regional response threshold for each assemblage. This taxa-specific and integrative thresholds are established for taxa that respond positively (increase in abundance) and negatively (decrease in abundance) to increasing nutrients. This makes TITAN an ideal technique to derive structural thresholds because it provides stressor 'bookends' with one threshold that identifies concentrations where the most sensitive taxa are lost and another that identifies conditions where tolerant taxa thrive (Kail et al. 1012, King et al. 2011).

TITAN measures changes in composition, but such compositional changes do not necessary imply degradation of aquatic life uses. Another approach to threshold derivation evaluates the nutrient concentrations that most closely correspond to independently derived measures of aquatic life degradation. The Utah Division of Water Quality (UDWQ) calculates macroinvertebrate O/E ratios for bioassessments (2014 Integrated Report). These estimates of condition are derived from analytical techniques that have been employed in numerous settings for over two decades. In brief, O/E ratios are derived from empirical models that compare the taxa expected (E) at a site without anthropogenic degradation against those predicted taxa that are actually observed (O). The E values are derived from empirical models that use compositional differences among regional reference sites to make site-specific predictions of expected taxa. To make these predictions, the models use site-specific measures of natural environmental gradients—those unlikely to respond to human-caused stressors (geology, geography, precipitation, etc.)—to predict the probability of capturing each of the taxa (i.e., genus or species) that are part of the regional species pool. Once created, these models can then derive biological expectations of other sites (i.e., those not used in model development) based on their site-specific physical and geographical characteristics. Further, the biological expectations, expressed as probability of capture, are provided for all known taxa within the modeled region (Hawkins et al. 2000). UDWQ recognizes that there are other methods of biological condition assessment and that the models described above are not exhaustive in their accounting for site-specific covariables. However, the models do account for broad physical and geographical variables and because they are used by UDWQ as quantitative estimates of the support of aquatic life uses, O/E can be used to estimate N or P thresholds that, on average, are associated with independently derived measures of biological degradation.

This chapter attempts to answer several important questions related to the derivation of nutrient criteria for Utah's streams:

1. What concentrations of N and P are associated with the largest changes in the distribution and abundance of *sensitive* macroinvertebrates and diatom taxa? How do these concentrations differ for tolerant taxa?
2. What concentrations of N and P best distinguish between biologically degraded and non-degraded streams? Are these thresholds dependent upon different analytical methods?
3. Do sites that exceed N or P thresholds correspond with sites that would be considered biologically degraded from independent biological assessments?

Methods

Stream Sites (Data Selection)

The diatom and macroinvertebrate data that we used for these analyses were collected from 370 streams across Utah between 2001 and 2010. We compiled these data from several sources, including: Utah's Comprehensive Assessment of Stream Ecosystems (UCASE), the National Wadeable Streams Assessment (WSA), National Rivers and Streams Assessment (NRSA), UDWQ specific programmatic sampling events (standards development), and the United States Geological Survey's (USGS) National Water-Quality Assessment (NAWQA) Program (Cuffney et al 2005). Field methods among programs followed nearly identical sampling procedures (USEPA 2007). Irrespective of the data source, water chemistry and habitat characteristics were collected immediately prior to the collection of biological samples.

While protocols were nearly identical among collection programs, there were several differences in laboratory methods that affected how we ultimately treated the data. One important difference was whether or not TN data were available. TN was only recently added to UDWQ's regular chemical analytical suite, which limited the number of sites that UDWQ was able to evaluate and report on in this chapter. Labs also differed with respect to their reporting limits for different parameters. Many samples came from UDWQ studies and the Utah State Health Laboratory where the samples were processed with a detection limit of 0.02 mg/L TP, while other samples used in these analyses (WSA, NRSA & NAWQA) had lower detection limits. In all cases, we used a value of one half of the lab-specific reporting limit for these analyses.

Biological Data Collections

Biological samples were obtained following USEPA's Rapid Bioassessment Protocols SOPs (Barbour et al. 1999). UDWQ's diatom bioassessment program is relatively new, so while all sites with diatoms also had corresponding macroinvertebrate data, the converse was not always true. This limitation coupled with limited TN data (see above) means that the number of sites differed for each analysis, because we applied the goal of always maximizing sample size for each analysis. Diatomswere collected from hard surface benthos—typically cobbles— from 11 transects at each stream. Diatom samples were chilled to preserve and then subsequently identified to lowest practical taxonomic resolution by Rushforth Phycology, LLC following standard laboratory procedures (Rushforth Phycology 2005). Macroinvertebrate data are based on a composite of 8 fixed-area riffle samples. Macroinvertebrates samples were preserved in ethanol and processed to lowest practical taxonomic resolution by the BLM BugLab using 500-ct fixed-count subsample methods (Miller and Judson 2011).

Analytical Methods

COMPOSITIONAL CHANGES (TITAN)

TITAN in R v2.15.0 (R Core Development Team 2012) was used from package `TITAN` and the analytical procedures described in Baker and King (2010); UDWQ had diatom and macroinvertebrate data for 370 sites throughout Utah. We subsequently screened these sites based on the availability of TN data, which left insufficient data to calculate nitrogen thresholds for diatoms. There were, however, 251 sites with diatoms and TP, which was more than sufficient for calculating TP thresholds from TITAN. Similarly, macroinvertebrate TITAN models were created based on samples obtained from 178 sites across Utah. For both assemblages, TN and TP were used independently as environmental stressors in TITAN analyses. In this exercise, an exhaustive list of site-specific covariables was not incorporated into the TITAN analyses. Site-specific covariables could be assessed, as appropriate, in development of site-specific criteria (See Chapter 12).

BIOLOGICAL IMPAIRMENTS

We used RIVPACS models that were created for the 2010 *Integrated Report* to generate O/E scores (UDWQ 2010, pp. 31-36 for details). Macroinvertebrate O/E scores were calculated with data collected from 243 stream sites. These sites include 97 reference sites and another 146 randomly selected sites that best represent the range of conditions found throughout the State (Olsen and Peck 2008). All sites were used to make O/E and TP comparisons, whereas our comparisons with TN were limited to 68 sites where both TN concentrations and O/E scores were available. We used untransformed TN and TP data to conduct TITAN analyses, but subsequently log transformed the data prior to post-hoc parametric statistical evaluations so that these data met the underlying statistical assumptions of these tests.

We evaluated relationships between O/E scores and the TN or TP data with simple linear regressions. Next, we followed UDWQ's established biological assessment procedures and categorized each O/E score into three categories based on the extent to which models were able to reliably detect departure from reference condition: Good (O/E >0.83; 5% Type I error rate), Fair (<0.83 but >0.78) and Poor (<0.78, 10% Type I error rate) (UDWQ 2010 , p. 35). We then evaluated relationships between Good and Poor sites and TN or TP concentrations (mg/L) with logistic regression; we dropped sites categorized as Fair (n=10) for this analysis because the logistic regression requires binary response data. We then developed nonparametric deviance reduction (NDR, Qian et al. 2003) models to identify thresholds of TN and TP that best distinguish between good and poor sites (as independently defined by O/E scores). Bootstrapping (10,000 replications, R package `boot`) was subsequently used to calculate 95% confidence intervals for each threshold. Finally, we used a two sample t-test with pooled variance scores to determine if O/E scores at above and below TN and TP NDR thresholds were statistically different ($p<0.05$).

FURTHER EVALUATIONS OF THRESHOLDS DERIVED FROM BIOLOGICAL ASSESSMENTS

We further evaluated the NDR nutrient thresholds with Receiver Operating Characteristic (ROC) and Relative Risk (RR) analyses. ROC analyses confirm the appropriateness of thresholds in the context of regulatory decisions (Morrison et al. 2003, Carlisle et al. 2009, McLaughlin 2012), whereas RR analyses evaluate the extent to which the thresholds identify the extent of risk to stream biota.

Receiver Operating Characteristics

ROC analysis (R package pROC) allowed us to identify thresholds that minimize false positive and false negative assessments—as defined by independently derived O/E impairment thresholds. ROC calculates a single value for model fit called area under the curve (AUC), which is the probability that a randomly chosen response above the threshold will be greater than a randomly chosen response below the threshold. This is approximately the same procedure used in many nonparametric ranked tests such as the Mann-Whitney U or the Wilcoxon rank sum test (Mason & Graham 2002). Error rate estimates (or non-error rates) are also provided by ROC and are generally more germane to this study than the overall model fit because they relate to decisions made in the biological assessment process: Is a stream supporting the biological uses or not?. The analytical basis of ROC is a 2x2 error matrix (or confusion matrix) representing two states of condition (impaired or not impaired) and two states of predicted condition (i.e., good vs. poor) in relation to a continuous stressor variable (McLaughlin 2012). Error rate statistics were calculated for the proportion of true positives (when nutrient threshold is exceeded and the site is biologically impaired) and the corresponding proportion of true negatives for the range of TN and TP stressors observed among sites. ROC was also used to predict optimal (as defined by the researcher) Type I (false positive) and Type II (false negative) error rates, to identify the stressor response thresholds that maximize overall model performance (Nevers and Whitman 2011, Hale and Helsthe 2008).

Relative Risk

We ran relative risk (RR) analyses to identify the threat to aquatic life for sites above and below nutrient thresholds with R package `spsurvey` (v2.2). RR analyses provide an estimate of the relative threat of TN and TP, along with the 95% confidence intervals (CIs) surrounding these estimates. RR analysis is commonly used in the medical field and its interpretation is straight forward: What is the factor by which risk increases following exposure to a stressor? Like ROC, a 2x2 error matrix underpins RR analyses. However, in RR the response and stressor variables must be both categorical and binary. In this case, predefined nutrient stressor thresholds (i.e. high and low) and the relationship to a binary biological response (i.e., poor or good condition) were examined. The subsequent results indicate the increased risk, relative to other risks evaluated, that the biota will be in poor condition if the stressor exceeds levels of greatest threat to stream biota (Van Sickle and Paulsen 2008). We considered a risk significant if the lower confidence interval was greater than 1.0 (Van Sickle et al 2006, Van Sickle & Paulsen 2008) for either N or P. All analyses were conducted in R v 2.15.0 (R Development Core Team, 2012).

Results

Compositional Changes

DIATOMS

Diatom compositional changes could only be evaluated for TP. TITAN revealed thresholds

Table 6.1. Community level threshold responses of diatoms to total phosphorus determined by TITAN significant responders and nonparametric change point analysis using Euclidean distance (nCPA).

Diatoms		Total Phosphorus (mg/L)		
Community	Method	Threshold	5th Percentile	95th Percentile
Sensitive	TITAN	0.016	0.010	0.022
Tolerant	TITAN	0.042	0.027	0.051
All	nCPA	0.022	0.010	0.047

for eight diatom taxa that significantly decreased in abundance and occurrence in response to increasing TP concentrations (Figure 6.1). Taxa were considered significant responders only if purity and reliability were > 0.95. The threshold response for all sensitive taxa—those that decreased with increasing TP— was at a TP of 0.016 mg/L (95% CI = 0.010-0.022 mg/L). Similarly, we found that, on average, the 29 diatom taxa that were statistically tolerant of increasing TP occurred at a TP of 0.042 mg/L TP (95% CI = 0.027-0.047 mg/L). Together, these two thresholds suggest that, on average, the diatom assemblage starts to show appreciable losses of sensitive taxa at a TP of 0.016, and that more tolerant diatom taxa start to dominate the assemblage at a TP of 0.042 mg/L. The overall threshold that captures the TP associated with the most appreciable changes in both tolerant and abundant diatom taxa was at 0.022 mg/L (nCPA, 95^{th} CI = 0.010-0.047 mg/L). All significant diatom taxon-specific thresholds can be found in Table 6.3.

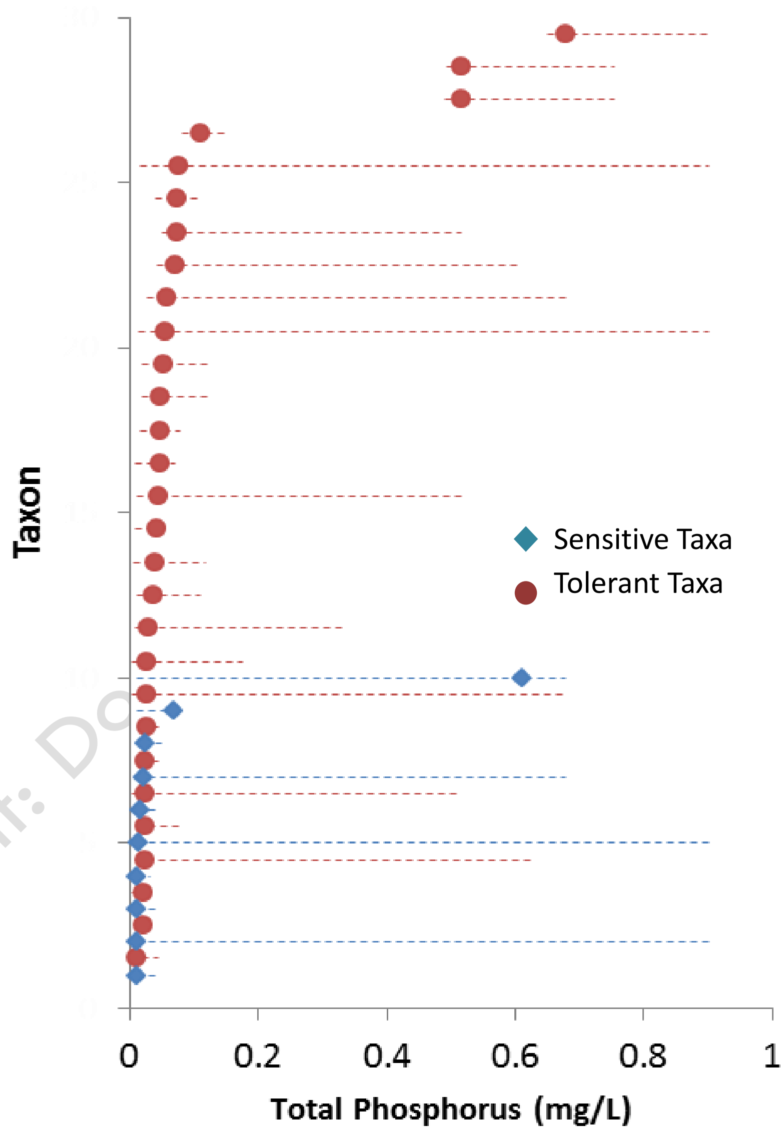


Figure 6.1. Significant indicator diatom taxa plotted in order of their environmental threshold as calculated by TITAN. Blue symbols represent sensitive (negative responders) taxa while red symbols represent tolerant (positive responders) taxa. Dashed lines indicate 5^{th} and 95^{th} percentiles determined by 500 bootstrap replicates.

MACROINVERTEBRATES

With the exception of tolerant taxa thresholds, the TP thresholds that we obtained from TITAN for macroinvertebrates were similar to those found for diatoms. We found significant ($p<0.05$, calculated from IndVal scores) thresholds for 47 sensitive macroinvertebrate taxa. These taxa significantly decreased in occurrence and abundance as among-site TP concentrations increased, which resulted in an assemblage level threshold for sensitive taxa at a TP of 0.011 mg/L (95% CI = 0.003-0.043, Figure 2B). Only 24 macroinvertebrate taxa were significantly tolerant of higher TP concentrations, and the most appreciable increases in occurrence and abundance occurring at a TP of 0.612 mg/L, which was appreciably higher than the threshold for tolerant diatom taxa. The overall assemblage-level shift from sensitive to tolerant taxa occurred at 0.015 mg/L TP (95% CI = 0.004-0.113mg/L) (Table 6.2).

Overall, TN resulted in fewer macroinvertebrate taxa with significant increases or decreases than TP. TITAN identified 40 sensitive macroinvertebrate taxa that significantly decreased in abundance and occurrence with increasing TN, whereas 17 significant tolerant taxa were identified (Figure 6.2). Additional work on of the extent to which temperature, shading, and substrate covary with ecological responses are discussed in Chapter 12 and will be applied during development of site-specific criteria. Significant individual taxa responses were accumulated into an assemblage-level response resulting in a TN threshold for sensitive macroinvertebrates of 0.18 mg/L TN (95% CI 0.14-0.40 mg/L). The TN threshold for tolerant macroinvertebrates was 0.41 mg/L TN (95% CI = 0.36-5.1 mg/L), and the assemblage-level shift from sensitive to tolerant taxa occurred at 0.41 mg/L (95% CI = 0.40-1.1 mg/L TN) (Table 6.2). All significant macroinvertebrate taxon-specific thresholds can be found in Table 6.3.

Biological Impairments

Macroinvertebrate O/E scores decreased with increasing nutrients. There was a significant, albeit weak, linear relationship among macroinvertebrate O/E scores and TN ($n=68$, $r^2 = 0.302$, $p<0.001$) and TP ($n=243$, $r^2 = 0.294$, $p<0.001$). The weakness in the relationship may be evidence

Table 6.2. Community level threshold responses of macroinvertebrates to total nitrogen and total phosphorus determined by TITAN significant responders and nonparametric changepoint analysis using Euclidean distance (nCPA).

Macroinvertebrates		Total Nitrogen(mg/L)			Total Phosphorus (mg/L)		
Community	Method	Threshold	5th Percentile	95th Percentile	Threshold	5th Percentile	95th Percentile
Sensitive	TITAN	0.18	0.14	0.40	0.011	0.003	0.043
Tolerant	TITAN	0.41	0.36	5.10	0.612	0.042	1.81
All	nCPA	0.41	0.40	0.1.1	0.015	0.004	0.113

of other stressors and natural gradients that are expected in structural responses because the effects of nutrients are indirect. For this reason, UDWQ has relied primarily on functional responses with more direct linkages to nutrients.

Macroinvertebrate scores were reorganized into binary data (impaired and not impaired) according to UDWQ's previously defined biological impairment classes to identify TN and TP concentrations associated with biological impairments. Logistic regression, models found both TN and TP concentrations to be significantly related to biological impairment indicator categories (odds ratio=2.27 [95% CI = 0.81-4.17], z=2.65, p=0.008, and odds ratio=44.51 [95% CI = 31.7-58.99], z=6.42, $p<0.001$, respectively).

Nonparametric deviance reduction (NDR) was used as a classification procedure, to determine specific thresholds in TN and TP concentrations that best differentiate 'Impaired' and 'Not Impaired' O/E conditions. Sites identified as 'impaired' based on macroinvertebrate O/E scores most

Figure 6.2. Significant indicator macroinvertebrate taxa plotted in order of their environmental thresholds to total nitrogen (A) and total phosphorus (B) as calculated by TITAN. Blue symbols represent sensitive (negative responders) taxa while red symbols represent tolerant (positive responders) taxa. Dashed lines indicate 5^{th} and 95^{th} percentiles determined by 500 bootstrap replicates.

frequently occurred at streams with TN >0.41 mg/L (n=68, 95% CI = 0.12-0.79 mg/L). This threshold correctly predicted 68 percent of true positives (prediction probability (PP) =0.68) and 76% of true negatives (PP=0.76). NDR also identified an average TP threshold of 0.045 mg/L (n=232, 95% CI = 0.023-0.066 mg/L), which was also quite accurate (true positive PP= 0.65 and true negative PP=0.89).

Significant differences were found among the O/E scores between the high and low TN sites ($t=4.22$, $p<0.001$, Figure 6.3A) and the high and low TP sites ($t=-3.88$, $p<0.001$, Figure 6.3B) using two-sample t-tests with pooled variances. The significant differences between O/E scores at the high and low TN and TP sites provide further support for the significance of these thresholds.

Alternative Statistical Methods

Lastly, the strength of our thresholds with ROC and RR were evaluated. ROC revealed that the TN stressor-response model predictions were quite accurate, with a 77.3% chance that randomly selected site below the TN threshold of 0.42 mg/L will have a higher O/E score than a site above the threshold (area under the curve (AUC) = 77.3 (95% CI =64.2-88.2)). The TP threshold of 0.045 mg/L performed even better with an AUC =81.4 (95% CI = 75.3-87.3). Next, we evaluated all possible TN and TP thresholds to identify nutrient concentrations that maximized the percent of both true positives and true negatives. These analyses revealed that the maximum number of both true positives and true negatives occurs at a TN of 0.33 mg/L TN and a TP of 0.045 mg/L TP (Figure 6.4). Although, depending on the management objective, balancing Type I and II errors may not always be appropriate. Development of site-specific criteria will incorporate additional analyses of potential covariables to reduce both Type I and Type II errors.

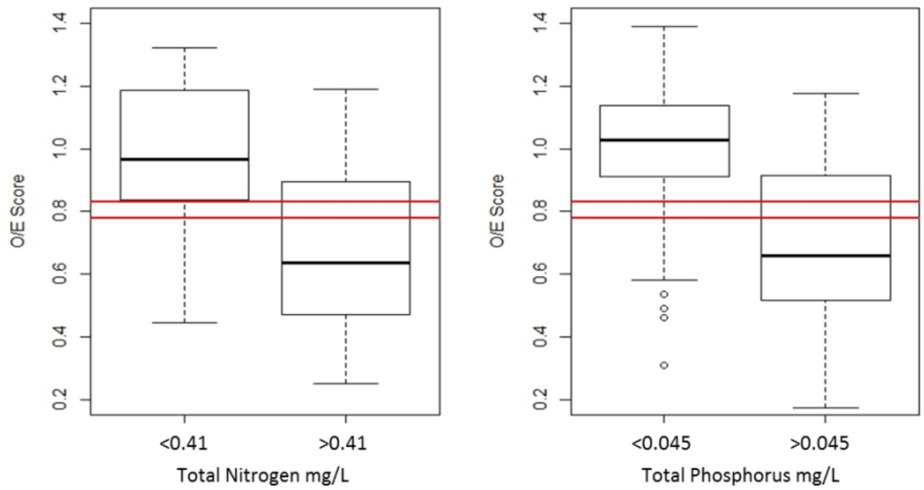


Figure 6.3. Box plots of numeric O/E scores for sites above and below the thresholds determined by nonparametric deviance reduction for total nitrogen (A) and total phosphorus (B). Thresholds were developed from categorical O/E scores between impaired and not impaired sites. The double red lines indicate the range of Fair O/E scores while above is considered Good (not impaired) and below is considered Poor (impaired).

Relative risk (RR) analyses provided another line of evidence that O/E impairments were associated with higher concentrations of both TN and TP. Using RR analysis, the nutrient thresholds established from NDR resulted in a RR for TN of 2.09 (95% CI = 1.57-2.95). This indicates that if the TN threshold is exceeded at a site there is a 2.09 fold greater chance that the O/E score will also indicate impaired conditions compared to a site below the TN threshold. Using the same analysis, an even stronger RR of 3.66 was found for the TP threshold (95% CI=2.66-5.01).

Discussion

Several lines of evidence explored the changes to various measures of macroinvertebrate and diatom assemblages at varying nutrient concentrations. Overall, we demonstrated that specific taxa respond consistently to changes in TN and TP concentrations. Also, by comparing TN and TP to biological assessments we demonstrate clear relationships between increasing nutrients and independently derived measures of biological condition. Finally, our risk and ROC analysis helped establish the TN and TP thresholds based on biological assessment metrics in the context of the likelihood of impairments and predicted Type 1 and Type II errors.

Despite the concordance among several lines of evidence, several sources of bias potentially limit the broad applicability of these results. First, these analyses were conducted using data from different State and Federal sources, which potentially biases results by excluding or over representing environmental gradients. However, such bias is unlikely given that the sites are broadly dispersed geographically and all of the physicochemical parameters that were evaluated covered the range of conditions observed among randomly selected statewide sites (UDWQ, WSA & NRSA). Nevertheless, the potential for site selection bias remains because UDWQ selects sites based on programmatic needs, which often either target reference sites (UDWQ, WSA & NRSA) or sites with known or suspected water quality problems. Ideally, a predetermined study design could be developed to better control covariates such as other stressors or natural environmental gradients. This lack of control with sample design may have decreased the strength of threshold-response relationships. Nonetheless, these analyses merely demonstrate a correlation of high nutrient concentrations with biological degradation. UDWQ recognizes that degradation of biological condition relates to stresors and variables other than nutrients. For example, because this was a state-wide exercise, covering large ranges in elevation, watershed size, watershed perturbations, stream size, local habitat conditions, etc., other regional and site-specific covariate stressors, such as temperature, substrate size, stream velocity, shading, etc. could play a significant role in nutrient threshold determination. Despite these shortcomings, these results, in conjunction with the other lines of evidence presented in this report support the axiom that excess nutrient inputs ultimately degrade stream biota.

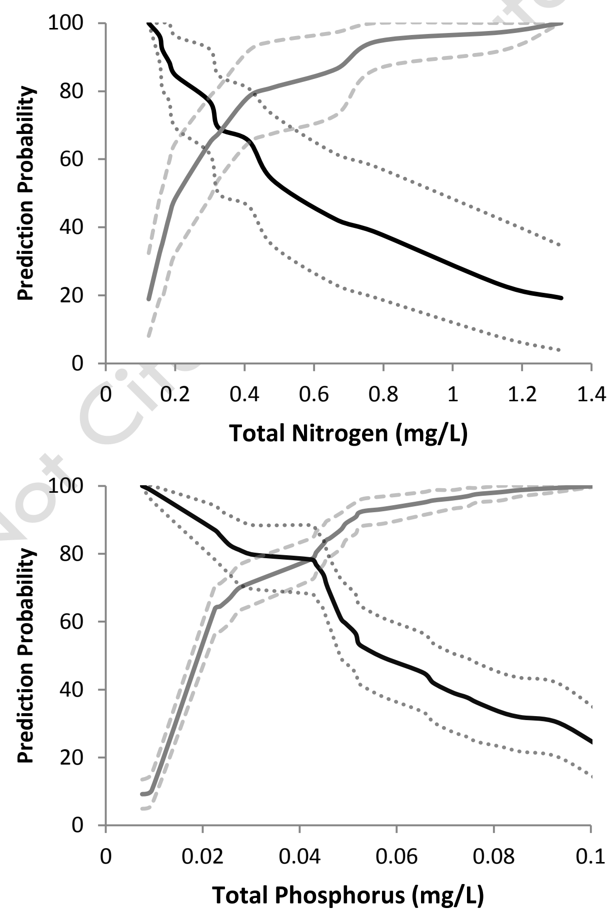


Figure 6.4. Prediction probability curves generated from ROC analysis. Grey line indicates the positive prediction probability (sensitivity) of correctly predicting a biologically impaired site (by O/E score) at a given numeric threshold. The black line indicates the negative prediction probability (specificity). Dashed lines are 95% confidence intervals from 2000 bootstrap replicates.

Changes in Stream Assemblage Composition and in Response to Nutrients

TITAN analysis is built on the estimation of taxon-specific change-points, which is arguably its greatest strength because it evaluates changes in biological structure fundamentally different than more traditional community-based measure of biological condition (IBI, O/E, etc.). As a result, TITAN models can be built to evaluate biological responses to any environmental stressor and where multiple stressors co-occur, it is important to evaluate each one to determine the most effective remediation practice. Not all sensitive taxa are equally sensitive to all stressors, so it is equally important that TITAN only considers predictable (significant) responders in the calculation of stressor-response thresholds and implementation planning will consider all stressors on a site-specific basis. Conversely, differential taxa responses are an intrinsic problem with community-based metrics because taxon-specific responses have the potential to be based in aggregate.

TITAN generated two thresholds developed for both TN and TP that together provide 'bookends', with a range of nutrient concentrations that are sufficient to support stream biota. However, UDWQ does not aim to protect _all_ of the most sensitive species. Even under completely undisturbed natural conditions it is likely that diatom and macroinvertebrate taxa have distributions that vary with physicochemical factors (i.e. floods, droughts, shading, temperature), some of which could either protect against or exacerbate nutrient responses. Conversely, allowing nutrient concentrations to achieve levels that result in assemblages dominated by tolerant taxa are likely underprotective because they fail to meet diversity, stability, and biodiversity objectives intrinsic to commonly accepted interpretations of biological integrity, a major CWA objective. TITAN provides the TN and TP concentrations that are, on average, associated with both of these extremes. A sufficiently protective criterion likely falls somewhere between these two benchmarks, so these brackets provide meaningful context.

Diatoms vs. Macroinvertebrates

Total phosphorus thresholds that were based on the most sensitive taxa were similar for diatoms and macroinvertebrates (0.016 and 0.011 mg/L, respectively). These concentrations are near our best estimates of background concentrations (Chapter 10), which implies that subtle assemblage-level shifts are initiated at fairly low concentrations. TP thresholds derived from tolerant diatoms remained low (0.022 mg/L), whereas thresholds for tolerant macroinvertebrates were appreciably higher (0.612 mg/L) than those derived from sensitive taxa. This divergence in assemblage responses is consistent with other investigations that have generally reported diatoms to be more sensitive to nutrient enrichment than macroinvertebrates (i.e., Hering et al. 2006), although among assemblage differences can sometimes be obscured by the specific biological metrics or analytical methods used to derive thresholds (Johnson et al. 2006).

Relationship to Biological Impairments

Simple linear regression found a significant, although weak, negative relationship between nutrients and expected macroinvertebrate communities (measured as O/E). Over a large range of nutrient concentrations, O/E scores are expected to have a negative relationship with nutrients like most stressors, yet this this relationship in unlikely linear. At very low nutrient concentrations small increases may actually increase community health by increasing primary and secondary productivity (Hart and Robinson 1990, Mazumder and Edmonson 2002, Slavik et al. 2004). At intermediate levels of nutrient concentration, the negative effects may be either masked or exacerbated by site-specific characteristics (shading, scouring, grazers, etc.) (Liess et al. 2009). These sources of variability, left unaccounted, necessarily obscure linear stressor-response relationships. Thus, while linear relationships are helpful to evaluate trends, non-linear and multivariate analyses are almost always required to elucidate nutrient thresholds that are truly detrimental to aquatic communities.

This chapter reported on a strong relationship between biological impairment and in-stream TN and TP concentrations as determined by logistic regression and NDR. The logistic regression models showed a much stronger fit for TP than TN (odds ratio= 44.5 and 2.3, respectively), although both were significant. The odds ratio can be interpreted as for every one unit of change in nutrient concentration the odds of having a corresponding impaired biological condition increases by 44.5% for TP and 2.3% for TN. The difference in magnitude of these odds ratios may be a factor of the TP sample size being much larger than TN (n=243 and 68). Whether TP really has a much stronger relationship to biological impairments still has some uncertainty. However, it is clear through this analysis that as nutrients increase the odds of degraded macoinvertebrate assemblage increases.

Nonparametric deviance reduction (NDR) enabled further expansion of logistic regression results; providing numeric TN and TP thresholds that best distinguish between sites. On average, O/E scores on either side of TN and TP thresholds significantly differed. NDR identified thresholds that were closer to the higher thresholds determined by TITAN for both TN and TP, which suggests that thresholds based on sensitive taxa responses may be overly protective. These indicators provide another objective metric that can be used to identify streams that are most likely to have nutrient-related impairments for follow-up site-specific confirmation.

Corroboration among Analytical Methods

ROC results confirmed the NDR-derived thresholds by demonstrating that randomly selected O/E scores indicated degraded condition for sites above TN and TP thresholds, in 77% and 81% of cases respectively. Perhaps the most interesting ROC insight is the ability to determine Type I and Type II errors at any given threshold (Figure 6.4). If the goal is to maximize model performance then an indicator threshold would be located at the intersection of true positive and true negative prediction probabilities, which equates to 0.33 mg/L TN and 0.045 mg/L TP. If the goal is to minimize Type I errors (false positive impairment conclusions), in order to maximize the chance that resource intensive site-specific follow-up investigations are focused on real environmental problems, then a threshold that maximized true positive predictions should be emphasized. Alternatively, the importance of false negatives could be emphasized (typically defacto $\alpha = 0.05$, $\beta = 0.20$) if resource managers wanted to err on the side of the Precautionary Principle (e.g. UNEP, 1992) and avoid situations where real environmental problems go undetected. In either case, ROC helps inform management decisions by elucidating the intrinsic tradeoffs involved with pragmatic resource limitations (wasted time and effort addressing false positive, Type II, assessment errors) and mandate to protect water quality (missing impaired waters, Type I, assessment errors).

The relative risk analysis also confirms the results from logistic regression and ROC demonstrating that increases in nutrients leads to increased probability of impairments. RR is slightly different than ROC as it analyzes the stressor as a binary variable instead of a continuous variable. TN and TP thresholds determined by NDR were used to convert nutrient concentrations into a binary variable of "Good" and "Poor" based on ambient TN and TP concentrations. The results were similar for logistic regression as TP had a stronger effect than TN, although the relative difference was much smaller. This indicates that not only do increased nutrients indicate an increased probability of biological impariment, but managers can actually develop a numeric threshold for nutrient concentrations that represent this probability (or risk) of biological degradation without ever knowing the underlying O/E scores.

Overall, these additional analyses suggest that nutrient thresholds are not merely a statistical artifact, but instead are based on significant relationships with changes in the composition of stream assemblages. Combining threshold analyses with additional model performance evaluations, such as ROC and RR, has several advantages: improved defensibility of nutrient thresholds, insight into the risk that excess nutrients pose to aquatic life, and the ability to make the most informed management decisions possible. These analyses can be repeated as biological assessment thresholds are developed for additional freshwater assemblages (i.e., fish) to paint a more comprehensive picture of the relative sensitivity of a more diverse range of stream community taxa. Differences and similarities in among assemblage responses to nutrient enrichment should provide a more complete understanding of impacts to structural aspects of stream condition.

Summary and Recommendations

Macroinvertebrate O/E ratios have been the backbone of UDWQ's biological assessment program since its adoption in 2008. Utah's aquatic life uses require the protection of fish (cold water (3A), warm water (3B) and non-game) and other organisms in their food chain. O/E ratios provide quantitative estimates of the extent to which human activities have caused local extinctions of macroinvertebrate taxa, which are a fundamental component of stream food webs. The challenge with using biological assessments is often not in determining if a site is biologically degraded, but the cause(s) of the degradation. The results presented in this chapter suggest that high nutrient concentrations are related to biological impairments (as measured as O/E ratios). Results also demonstrate that thresholds can be developed that identify stream nutrient concentrations that best predict—with known risks and error rates—where detrimental effects to stream biota are most likely to occur. Alone these thresholds could not predict nutrient-related impairments, but when coupled with other indicators (e.g functional responses) biological assessment data can be used to demonstrate that impairments are caused, at least in part, by excessive nutrient inputs.

The regionally derived, structural- based thresholds reported in this chapter describe N or P concentrations that, on average, are associated with alterations in the composition of stream biota. Such regional indicators are useful because they can more accurately identify sites with potential nutrient-related problems. However, management decisions are ultimately applied to specific watersheds. Hence, once sites with nutrient-related problems are identified, additional site-specific investigations will need to be designed to more carefully examine cause-effect relationships between nutrients and stream biota. Specifically studies will need to elucidate the relative role of nutrients and other stressors to the loss of stream biota (Chapter 12). These investigations will also need to determine how local habitat conditions (i.e., covariates) diminish or exacerbate the effects of nutrient enrichment on macroinvertebrates or diatoms. UDWQ has incorporated tiered monitoring and assessment approaches to accommodate the transition from regional trends to site-specific conditions.

Table 6.3 Taxon-specific sensitivity values from TITAN models. Changepoint is the threshold of TP (mg/L). Maxgrp 1 indicates sensitive (negative responders) taxa and maxpgrp 2 indicates tolerant taxa (positive responders). There are two important diagnostic indicators calculated from 500 bootstrap replicates. Purity is the proportion of changepoint response direction that corresponds with the observed response direction. Reliability is the proportion of changepoints with an IndVal score that results in significant p-values. Z-scores are calculated by standardizing IndVal scores by subtracting its permutated mean and dividing by its permutated standard deviation. Z-scores may be a better metric than IndVal scores when comparing strength of response between widely distributed taxa and rare taxa.

Part 1: Sensitivity of Diatom Taxa to TP

Diatom-TP (mg/L)										
Taxon	Changepoint	freq	maxgrp	IndVal	pval	z score	5th Percentile	95th Percentile	Purity	Reliability
Anomoeoneis vitrea (Grunow) Ross	0.010	10	1	5.66	0.024	2.56	0.010	0.044	0.96	0.78
Diploneis oblongella (Naegeli) Cleve-Euler	0.010	18	1	10.11	0.008	3.92	0.010	0.901	0.74	0.73
Meridion circulare (Greville) C.A. Agardh	0.010	25	1	13.46	0.004	5.08	0.010	0.043	1.00	0.95
Cymbella naviculaformis Auersw. Ex Heib.	0.010	10	1	5.96	0.016	2.57	0.010	0.032	0.97	0.85
Synedra delicatissima W. Smith	0.013	10	1	6.92	0.004	4.03	0.010	0.901	0.77	0.74
Hannaea arcus (Ehr.) Patrick	0.015	17	1	7.87	0.02	2.67	0.010	0.045	0.86	0.76
Cymbella affinis Kützing	0.021	152	1	37.38	0.044	2.1	0.010	0.679	0.95	0.87
Achnanthes minutissima Kützing (Achnanthidium)	0.024	247	1	55.82	0.004	8.77	0.022	0.059	1.00	1.00
Cymbella microcephala Grunow	0.070	59	1	22.28	0.008	3.24	0.010	0.080	0.98	0.94
Cymbella minuta Hilse ex Rabenhorst (Encyonema)	0.609	157	1	60.41	0.004	3.55	0.010	0.679	0.99	0.98
Amphora perpusilla Grunow	0.010	199	2	52.67	0.004	6.51	0.010	0.045	1.00	1.00
Nitzschia inconspicua Grunow	0.021	172	2	59.34	0.004	12.57	0.010	0.030	1.00	1.00
Gomphonema clevei Fricke	0.022	39	2	18.02	0.004	5.63	0.018	0.034	0.99	0.98
Achnanthes lanceolata (Breb.) Grunow	0.023	120	2	37.32	0.004	4.8	0.015	0.629	0.98	0.98
Gomphonema parvulum Kützing	0.023	137	2	40.89	0.004	5.46	0.012	0.082	0.98	0.98
Surirella ovalis Brebisson	0.025	99	2	32.2	0.004	4.83	0.021	0.516	0.98	0.98
Nitzschia palea (Kützing) W. Smith	0.025	166	2	47.29	0.004	6.14	0.010	0.049	0.99	0.99
Navicula lanceolata (Agardh) Ehrenberg	0.026	85	2	30.74	0.004	5.63	0.010	0.049	0.99	0.99
Nitzschia tryblionella *Hantzsch*	0.026	14	2	8.39	0.004	4.1	0.023	0.679	0.97	0.95
Nitzschia hungarica Grunow	0.028	18	2	11.57	0.004	5.51	0.023	0.176	0.99	0.98
Surirella ovata Kützing	0.029	20	2	12.09	0.004	5.23	0.021	0.335	0.99	0.98
Cyclotella meneghiniana Kützing	0.038	64	2	34.93	0.004	10.29	0.026	0.111	1.00	1.00
Pinnularia species	0.040	8	2	9.36	0.004	6.89	0.035	0.117	0.99	0.98
Synedra ulna (Nitzsch.) Ehr.	0.041	155	2	51.67	0.004	7.83	0.034	0.047	0.98	0.98
Nitzschia sigmoidea (Nitzsch) W. Smith	0.044	11	2	11.56	0.004	7.18	0.034	0.516	1.00	0.99
Pinnularia brebissonii (Kutz.) Rabenhors	0.047	6	2	10.34	0.004	8.64	0.040	0.071	1.00	0.98
Cyclotella species	0.047	12	2	14.24	0.004	7.79	0.030	0.080	0.99	0.97
Amphora coffeaeformis (Agardh) Kützing	0.047	26	2	18.87	0.004	7.02	0.030	0.124	0.96	0.96
Synedra ulna var. *constricta* Venkt.	0.053	11	2	14.03	0.004	8.34	0.035	0.124	1.00	0.99
Fragilaria brevistriata Grunow (Pseudostaurosira)	0.055	21	2	20.65	0.004	9.05	0.041	0.901	1.00	1.00
Bacillaria paradoxa Gmelin	0.057	20	2	20.05	0.004	9.24	0.030	0.679	1.00	1.00
Navicula pygmaea Kützing	0.072	9	2	11.35	0.004	5.61	0.030	0.609	0.97	0.96
Navicula capitata Ehrenberg (Hippodonta)	0.075	25	2	20.45	0.004	6.57	0.025	0.516	0.99	0.99
Cymatopleura elliptica (Brebisson) W. Smith	0.075	9	2	14.84	0.004	7.51	0.034	0.111	0.97	0.95
Navicula minuscula Grun.	0.075	28	2	29.57	0.004	9.54	0.059	0.901	1.00	1.00
Nitzschia paleacea (Grunow) Grunow in van Heurck	0.111	136	2	49.82	0.008	4.97	0.030	0.150	1.00	0.99
Nitzschia valdecostata (Lange-Bertalot) Seimonson	0.516	16	2	30.63	0.004	7.36	0.025	0.755	1.00	1.00
Navicula tripunctata var. *schizomenoides* (Van Heurck) Patrick	0.516	40	2	40.28	0.008	5.58	0.023	0.755	0.99	0.98
Nitzschia apiculata (Gregory) Grunow	0.679	30	2	51.22	0.004	6.74	0.030	0.901	0.98	0.97

Part 2: Sensitivity of Macroinvertebrate Taxa to TN

Macroinvertebrate-TN (mg/L)											
Taxon		Changepoint	freq	maxgrp	IndVal	pval	z score	5th Percentile	95th Percentile	Purity	Reliability
EMPIDIDAE		0.068	57	1	76.29	0.004	7.28	0.052	0.382	1.00	1.00
NEMOURIDAE	*AMPHINEMURA*	0.068	14	1	39.22	0.004	7.37	0.051	0.323	1.00	0.98
DRYOPIDAE	*HELICHUS*	0.091	15	1	39.66	0.004	9.23	0.051	0.219	1.00	1.00
LEPIDOSTOMATIDAE	*LEPIDOSTOMA*	0.124	50	1	35.73	0.004	4.48	0.119	0.582	0.99	0.98
BRACHYCENTRIDAE	*MICRASEMA*	0.138	30	1	29.33	0.004	5.86	0.125	0.505	1.00	1.00
UENOIDAE	*NEOTHREMMA*	0.140	10	1	15.1	0.004	5.22	0.124	0.422	1.00	0.99
HYGROBATIDAE		0.142	47	1	27.4	0.008	3.43	0.091	1.085	0.99	0.98
EMPIDIDAE	*OREOGETON*	0.142	5	1	15.15	0.004	9.54	0.091	0.177	1.00	0.98
HEPTAGENIIDAE	*CINYGMULA*	0.142	26	1	25.18	0.004	5.88	0.125	0.792	1.00	1.00
HYDROPTILIDAE		0.142	43	1	31.26	0.004	5.71	0.091	1.011	0.98	0.98
PHILOPOTAMIDAE	*DOLOPHILODES*	0.142	9	1	13.51	0.004	5.2	0.091	0.514	1.00	0.98
RHYACOPHILIDAE	*RHYACOPHILA*	0.142	47	1	30.57	0.004	4.48	0.130	0.792	0.99	0.97
CHLOROPERLIDAE	*SUWALLIA*	0.151	8	1	14.58	0.004	7.48	0.091	0.219	0.99	0.97
APATANIIDAE	*APATANIA*	0.151	12	1	21.14	0.004	8.69	0.132	0.374	1.00	1.00
EPHEMERELLIDAE	*SERRATELLA*	0.151	14	1	18.57	0.004	5.78	0.138	0.504	1.00	1.00
LEPTOPHLEBIIDAE	*PARALEPTOPHLEBIA*	0.163	38	1	38.58	0.004	9.03	0.111	0.398	1.00	1.00
HEPTAGENIIDAE	*RHITHROGENA*	0.165	35	1	24.61	0.004	5.15	0.124	1.050	0.99	0.96
CHLOROPERLIDAE	*SWELTSA*	0.165	29	1	36.49	0.004	10.75	0.110	0.276	1.00	1.00
LEUCTRIDAE		0.166	9	1	15.84	0.004	7.87	0.109	0.365	1.00	0.99
DIXIDAE	*DIXA*	0.171	5	1	11.9	0.004	7.33	0.091	0.186	1.00	0.97
TIPULIDAE	*HEXATOMA*	0.177	35	1	30.46	0.004	6.9	0.151	0.333	1.00	0.99
BAETIDAE	*BAETIS*	0.179	144	1	53.8	0.012	4.1	0.124	2.874	1.00	0.99
AMELETIDAE	*AMELETUS*	0.184	25	1	27.2	0.004	7.85	0.140	0.458	1.00	1.00
GLOSSOSOMATIDAE	*GLOSSOSOMA*	0.184	17	1	23.21	0.004	8.07	0.140	0.249	1.00	1.00
CERATOPOGONIDAE		0.223	81	1	38.48	0.004	4.42	0.051	0.424	0.96	0.96
CERATOPOGONIDAE	*DASYHELEA*	0.228	13	1	15.66	0.004	7.08	0.052	0.252	1.00	1.00
HEPTAGENIIDAE		0.250	50	1	32.89	0.004	7.33	0.111	1.085	1.00	1.00
ELMIDAE	*HETERLIMNIUS*	0.276	27	1	27.35	0.004	8.7	0.132	0.458	1.00	1.00
CHIRONOMIDAE	*MICROPSECTRA*	0.283	141	1	54.41	0.004	5.99	0.165	0.454	0.99	0.99
SIMULIIDAE	*SIMULIUM*	0.298	115	1	49.3	0.004	6.68	0.094	0.933	1.00	1.00
EPHEMERELLIDAE	*DRUNELLA*	0.333	39	1	32.46	0.004	9.16	0.158	0.560	1.00	1.00
PERLODIDAE		0.387	53	1	28.9	0.004	4.44	0.080	0.772	0.99	0.97
CAMBARIDAE		0.422	10	1	10.42	0.004	4.68	0.119	0.483	1.00	0.98
CHLOROPERLIDAE		0.422	37	1	23.63	0.004	4.75	0.115	0.560	0.97	0.97
HEPTAGENIIDAE	*EPEORUS*	0.454	31	1	27.87	0.004	8.55	0.145	0.504	1.00	1.00
PERLIDAE	*HESPEROPERLA*	0.505	25	1	18.72	0.008	4.32	0.122	0.792	1.00	0.99
BRACHYCENTRIDAE	*BRACHYCENTRUS*	0.664	39	1	26.69	0.004	5.38	0.081	0.846	1.00	1.00
NEMOURIDAE	*ZAPADA*	0.702	39	1	28.4	0.004	6.23	0.163	0.881	1.00	1.00
ASELLIDAE	*CAECIDOTEA*	0.242	29	2	24.22	0.004	6.71	0.211	0.818	1.00	1.00
GAMMARIDAE	*GAMMARUS*	0.256	12	2	11.21	0.004	4.18	0.231	3.186	1.00	0.98
SIMULIIDAE		0.382	40	2	26.6	0.004	5.78	0.249	1.272	0.96	0.96
LEPTOHYPHIDAE		0.387	11	2	10.89	0.004	4.51	0.283	0.792	1.00	0.97
CHIRONOMIDAE	*EUKIEFFERIELLA*	0.398	175	2	60.13	0.004	5.1	0.233	0.910	0.98	0.98
HYDROBIIDAE		0.404	21	2	20.14	0.004	6.93	0.276	1.029	1.00	1.00
Hydrobiidae		0.404	38	2	29.74	0.004	7.43	0.298	1.500	1.00	1.00
OLIGOCHAETA		0.406	144	2	59.94	0.004	7.86	0.318	1.337	1.00	1.00
ERPOBDELLIDAE		0.504	26	2	26.55	0.004	9.61	0.357	1.067	1.00	1.00
EPHYDRIDAE		0.609	10	2	11.41	0.004	5.15	0.333	1.011	1.00	0.98
CORBICULIDAE	*CORBICULA*	0.664	7	2	9.32	0.008	4.39	0.382	1.848	1.00	0.97
PHYSIDAE	*PHYSA*	1.536	51	2	54.02	0.004	8.97	0.404	2.874	1.00	1.00
NEMATODA		1.562	61	2	49.75	0.004	7.08	0.256	3.186	1.00	1.00
HYALELLIDAE	*HYALELLA*	2.151	28	2	41.53	0.004	6.32	0.365	3.481	1.00	0.99
CORIXIDAE		2.430	12	2	34.49	0.004	8.31	0.541	3.481	1.00	0.99
PLANARIIDAE	*POLYCELIS*	3.481	41	2	48.84	0.008	4.1	0.221	5.079	0.98	0.97
COENAGRIONIDAE		3.511	45	2	58.59	0.004	5.01	0.177	5.079	0.97	0.95

Part 3: Sensitivity of Macroinvertebreates to TP

Macroinvertebrate-TP (mg/L)											
Taxon		Changepoint	freq	maxgrp	IndVal	pval	z score	5th Percentile	95th Percentile	Purity	Reliability
HYGROBATIDAE		0.002	47	1	50.51	0.012	3.61	0.002	0.197	0.976	0.910
LEUCTRIDAE		0.002	9	1	59.59	0.004	15.11	0.002	0.011	1.000	1.000
EMPIDIDAE	*OREOGETON*	0.003	5	1	35.79	0.004	11.69	0.002	0.010	1.000	0.986
OSTRACODA		0.003	34	1	63.22	0.004	9.83	0.002	0.009	0.974	0.974
CHLOROPERLIDAE		0.004	37	1	33.27	0.004	4.49	0.002	0.070	0.964	0.940
TAENIOPTERYGIDAE		0.004	9	1	36.28	0.004	12.92	0.002	0.008	0.988	0.988
CHIRONOMIDAE	*THIENEMANNIMYIA*	0.006	74	1	49.13	0.004	5.38	0.003	0.053	0.984	0.972
PSYCHODIDAE		0.006	26	1	37.52	0.004	8.14	0.002	0.011	0.998	0.998
PHILOPOTAMIDAE	*DOLOPHILODES*	0.006	9	1	22.83	0.004	9.77	0.002	0.011	1.000	1.000
PERLODIDAE	*MEGARCYS*	0.006	7	1	26.59	0.004	11.97	0.002	0.015	1.000	0.998
APATANIIDAE	*APATANIA*	0.007	12	1	50	0.004	17.79	0.002	0.009	1.000	1.000
HEPTAGENIIDAE	*RHITHROGENA*	0.007	35	1	43.23	0.004	8.81	0.002	0.017	1.000	1.000
ELMIDAE	*NARPUS*	0.009	10	1	16.83	0.004	5.62	0.002	0.028	0.988	0.950
NEMOURIDAE	*ZAPADA*	0.009	39	1	49.27	0.004	11.03	0.005	0.016	1.000	1.000
EPHEMERELLIDAE	*SERRATELLA*	0.010	14	1	27.33	0.004	9.51	0.002	0.015	1.000	1.000
HEPTAGENIIDAE	*CINYGMULA*	0.011	26	1	41.81	0.004	13.03	0.003	0.016	1.000	1.000
HEPTAGENIIDAE	*EPEORUS*	0.011	31	1	40.94	0.004	10.7	0.003	0.051	1.000	1.000
AMELETIDAE	*AMELETUS*	0.011	25	1	49.31	0.004	14.78	0.003	0.015	1.000	1.000
PERLIDAE	*HESPEROPERLA*	0.011	25	1	35.47	0.004	9.46	0.003	0.026	1.000	1.000
ELMIDAE	*HETERLIMNIUS*	0.011	27	1	38.2	0.004	10.44	0.002	0.036	1.000	1.000
CAPNIIDAE		0.011	23	1	23.02	0.004	5.92	0.002	0.018	0.966	0.960
LEPTOPHLEBIIDAE	*PARALEPTOPHLEBIA*	0.011	38	1	34.3	0.004	7.05	0.002	0.020	0.992	0.990
UENOIDAE	*NEOTHREMMA*	0.011	10	1	23.53	0.004	10.31	0.002	0.015	0.998	0.998
HYDROPSYCHIDAE	*PARAPSYCHE*	0.012	6	1	12.4	0.004	6.38	0.003	0.016	0.994	0.952
EPHEMERELLIDAE	*DRUNELLA*	0.014	39	1	44.88	0.004	11.66	0.003	0.029	1.000	1.000
RHYACOPHILIDAE	*RHYACOPHILA*	0.014	47	1	45.59	0.004	11.22	0.004	0.024	1.000	1.000
BRACHYCENTRIDAE	*MICRASEMA*	0.015	30	1	33.06	0.004	8.78	0.009	0.020	1.000	1.000
GLOSSOSOMATIDAE	*GLOSSOSOMA*	0.015	17	1	27.86	0.004	8.59	0.003	0.017	0.998	0.998
CHLOROPERLIDAE	*SWELTSA*	0.015	29	1	43.45	0.004	13.97	0.003	0.021	1.000	1.000
SPERCHONIDAE		0.017	80	1	43.36	0.004	5.89	0.004	0.022	0.980	0.978
ELMIDAE	*CLEPTELMIS*	0.017	33	1	28.66	0.004	7.1	0.007	0.036	0.978	0.970
HYDROPSYCHIDAE	*ARCTOPSYCHE*	0.019	30	1	31.25	0.004	9.45	0.006	0.031	1.000	1.000
CHLOROPERLIDAE	*SUWALLIA*	0.020	8	1	13.56	0.004	7.31	0.003	0.024	1.000	0.996
LEPIDOSTOMATIDAE	*LEPIDOSTOMA*	0.022	50	1	36.05	0.004	8.45	0.004	0.044	1.000	1.000
PTERONARCYIDAE	*PTERONARCELLA*	0.022	29	1	17.82	0.012	3.3	0.004	0.108	0.978	0.926
TIPULIDAE	*HEXATOMA*	0.024	35	1	23.58	0.004	5.57	0.002	0.027	1.000	0.986
CHIRONOMIDAE	*DIAMESA*	0.041	55	1	31.92	0.004	5.62	0.015	0.112	1.000	0.998
BRACHYCENTRIDAE	*BRACHYCENTRUS*	0.055	39	1	21.97	0.004	4.01	0.003	0.134	1.000	0.990
BAETIDAE	*BAETIS*	0.055	144	1	57.5	0.004	7.68	0.018	0.137	1.000	1.000
ELMIDAE	*ZAITZEVIA*	0.086	25	1	19.53	0.004	4.45	0.015	0.095	1.000	1.000
EPHEMERELLIDAE	*EPHEMERELLA*	0.097	32	1	21.32	0.008	3.81	0.003	0.108	0.996	0.986
HYDROBIIDAE		0.016	38	2	25.92	0.004	4.7	0.012	0.076	0.996	0.996
HALIPLIDAE	*BRYCHIUS*	0.017	29	2	21.06	0.004	4.75	0.015	0.074	0.990	0.986
EPHYDRIDAE		0.036	10	2	8.49	0.004	3.5	0.020	0.099	0.988	0.914
Erpobdellidae		0.039	26	2	16.41	0.008	3.54	0.009	0.099	0.986	0.948
LEPTOHYPHIDAE	*TRICORYTHODES*	0.063	55	2	31.44	0.004	5.02	0.008	0.273	1.000	0.996
OLIGOCHAETA		0.089	144	2	51.79	0.004	3.8	0.003	0.374	0.980	0.966
DOLICHOPODIDAE		0.099	8	2	15.12	0.004	6.89	0.082	0.341	1.000	0.984
ELMIDAE	*MICROCYLLOEPUS*	0.211	37	2	37.25	0.004	6.4	0.036	0.444	0.994	0.986
PLANORBIDAE	*GYRAULUS*	0.341	6	2	17.8	0.004	7.04	0.076	1.801	0.968	0.938
PLANORBIDAE		0.576	9	2	35.62	0.004	7.7	0.027	1.801	0.944	0.914
CORIXIDAE		0.700	12	2	43.35	0.004	7.88	0.033	1.801	0.964	0.934
PHYSIDAE	*PHYSA*	0.803	51	2	63.76	0.004	5.72	0.009	1.273	1.000	0.998
HYALELLIDAE	*HYALELLA*	1.273	28	2	54.56	0.004	6.29	0.019	1.801	0.998	0.978
COENAGRIONIDAE		1.273	45	2	70.15	0.004	6.2	0.025	1.801	0.990	0.956

CHAPTER 7

NUISANCE ALGAE AND RECREATION USE SUPPORT

Introduction

Most studies that have evaluated the effects of cultural eutrophication have focused on impacts to aquatic life uses, but impacts to recreation uses are well documented. Nutrient effects on recreation use are related to the quality, safety, and frequency of recreational use through two key mechanisms. Eutrophication related to nutrient loading is associated with algal overgrowth, which can reduce water clarity (turbidity) and color and increase growth of algal mats (periphyton) both of which reduce the frequency of recreation uses (Figure 6.1). Primarily in lakes, but also in some large rivers, cyanobacteria (blue-green algae) represent an additional threat to recreation uses.

Periodic overgrowth of algae violates the narrative water quality standard established by the State of Utah, which requires waters to be maintained such that they do not become offensive by "unnatural deposits, floating debris, oil, scum, or other nuisances such as color, odor, or taste;…or result in concentrations or combinations of substances which produce undesirable human health effects…" (UAC R317-2). The narrative standard established by the State of Utah already provides a regulatory basis for evaluating the influence of nutrient enrichment on recreation uses, however quantitative estimates of the algae concentrations where this occurs would simplify assessment processes.

Quantifying the extent to which nutrients have caused departure from natural or undisturbed conditions is relatively straight forward, whereas determining the point at which these changes are sufficiently appreciable to constitute a degradation of designated uses is more difficult. Such impairment decisions are based on resource management policies. Such resource management decisions can be informed by science, but science alone is insufficient because these decisions also depend upon societal values, which are sometimes captured in rules and regulations. While the nexus of science and policy is always challenging, this is particularly true when establishing criteria that are based on degraded aesthetics. Most people have experienced circumstances where degraded conditions have adversely affected our recreation experiences, but the conditions that cause these reactions depend on the values and experience of individuals. As the idiom states, "beauty lies in the eyes of the beholder".

Once waterbodies become sufficiently degraded people sometimes choose to recreate elsewhere (or not at all), which can have meaningful economic impacts to local communities. For instance, a study conducted in 2006 (Hoagland et al.) on the economic impacts of eutrophication determined that harmful algal blooms (HABs) in Maryland's coastal waters results in losses of $4 million per year due to recreation and tourism impacts. Another study found that direct losses of $10 million to Texas' fishing and tourism industry from a single HAB event (Evans and Jones 2001). Other problems occur when algae blooms affect the taste, odor, or color, so recreation impacts from algae blooms are not entirely the result of toxicity. UDWQ recently completed a related study and found that degraded water quality—primarily associated with excess nutrients and water clarity—causes annual losses of $20 million (CH2MHill 2012).

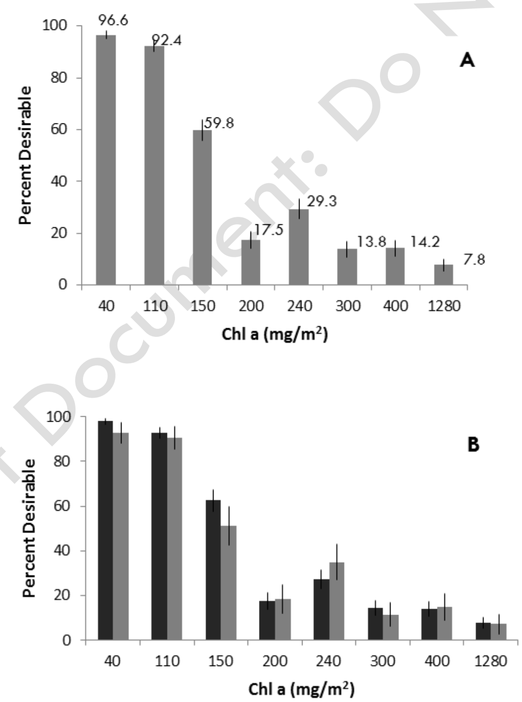


Fig 7.1. A) Percent desirable benthic algae response from all Utah survey participants. B) Percent desirable benthic algae responses from users (black) and non- user (grey) groups showing similarity in responses. Error bars indicate 95% confidence interval.

Recreation uses are explicitly or implicitly protected for all Utah streams and rivers. Currently these uses are protected with numeric criteria for *E. coli* to avoid harm to human health from fecal contamination (UAC R317-2). Aesthetics are also currently protected with narrative criteria (i.e., preclusion of scum, nuisance taste or odor, undesirable aquatic life), which applies to all of the state's surface waters. However, these statements are general and to date have primarily been used to preclude illegal dumping and littering. Given the importance of stream aesthetics, both to our quality of life and our economy, it remains important to better define conditions that constitute degradation of recreation uses. This chapter describes the results of an opinion survey sent to Utah households to determine what algae concentrations, if any, constitute undesirable recreation conditions.

Methods

UDWQ conducted a public opinion survey to determine whether excessive algae growth alters perceptions of aesthetics and desirable or undesirable recreation conditions. This survey was part of a larger research effort aimed at quantifying the economic benefits of avoiding cultural eutrophication in Utah's waters (see Ch2MHill 2012 for details). The surveys used for these analyses were mailed to 2,700 randomly selected Utah households, and we received 628 responses.

Survey participants were provided eight color photographs of streams with visible stream bottoms and varying benthic algae cover (Figure 7.2). For each stream in the photographs, Montana DEQ had previously quantified the extent of algae growth with per area *chl-a* concentrations that were obtained from a composite of 10-20 replicate algae samples (Suplee et al .2009). The benthic chlorophyll-*a* among these streams from <50 mg *chl-a*/m^2 to 1,276 mg *chl-a* /m^2 in intervals of~50mg/m^2 (Suplee et al. 2009). These data allowed UDWQ to quantify differences among sites, however the *chl-a* data were not sent to survey participants, because we wanted to ensure that their opinions were entirely based on an aesthetic response to the stream conditions depicted in the photographs.

Survey participants were shown the eight 8 photographs (Figure 7.2) in a randomly determined order. For each photograph participants were directed to "Please review the photos of algae in rivers on both sides of the one-page insert included in this survey. For each photograph on the insert tell us if the level of algae would be desirable or undesirable for YOUR most common uses of rivers, if any. There are no correct answers; this is your opinion only."

Table 7.1. Percent desirable survey responses among two user groups (user & non user) from Utah's survey and the Montana survey (Suplee et al 2009). Responses among users and between the two states did not differ (ANOVA p=0.94).

	Chl a	% Desirable Utah		% Desirable Montana	
Photo #	(mg/m^2)	Non-Users	Users	Non-Users	Users
1	40	92.8	98.0	95.6	98.2
7	110	90.6	92.9	94.9	93.6
6	150	51.2	62.5	69.7	75.8
5	200	18.3	17.4	16.5	31.8
2	240	34.9	27.1	28.8	29.1
8	300	11.4	14.4	12.6	20.2
3	400	14.7	14.1	16.7	11.5
4	1280	7.1	7.8	11.3	9.1

Survey responses were subsequently compiled and related to the quantitative measures of algae cover. We analyzed the relationship among benthic chl-*a* concentrations and percent desirable condition with a Spearman rank correlation. We also evaluated differences between user (those who recreate at stream) and non-user responses with a Kruskal-Wallis rank-sum test (both were considered significant at $p<0.05$). All analyses were conducted in R (R Development Core Team, 2012).

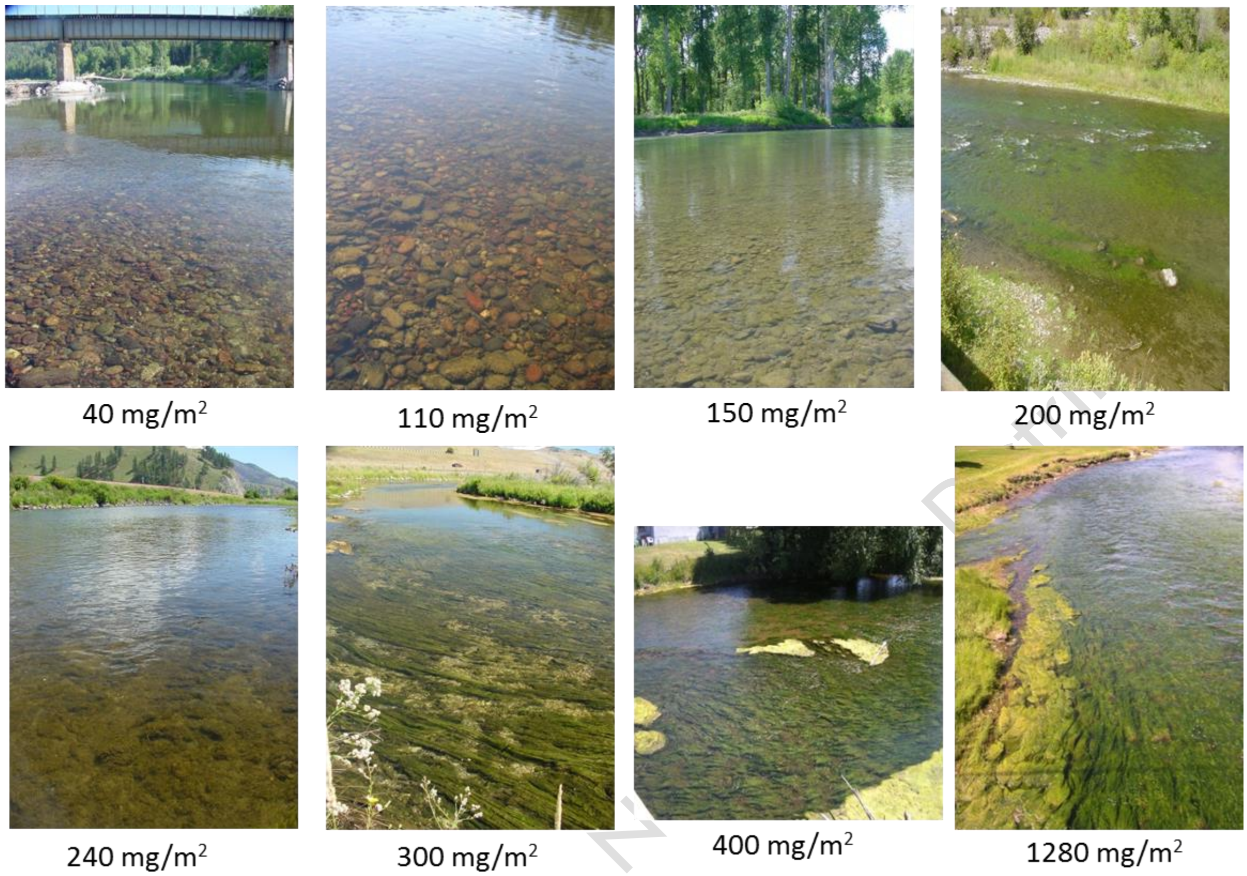


Figure 7.2. Photographs, obtained from Suplee 2009, that were provided to survey participants and associated chlorophyll a concentrations (not provided to survey participants.

Results

The survey revealed a strong and consistent relationship between benthic chl-*a* concentrations and what the public viewed as desirable conditions (Figure 7.1 A). The percent of respondents who indicated desirable conditions decreased from 96.6% at streams with benthic chlorophyll concentrations of40 mg/m^2 to a low of 7.8% at streams with 1280 mg/m^2 (Table 7.1). We found a significant negative correlation among benthic chl-*a* concentrations and percent desirable responses (Spearman's $r=-0.95$, $p<0.001$). We found two distinct thresholds among survey respondents. As benthic algae increased from 110 to 150 mg/m^2 the percent of "desireable" responses declined from 91 to 51%. The second threshold occurred at the next incremental increase in algae cover (110 mg/m^2) where desirable responses fell from 51% to 18%, and then remained consistently low for all subsequent incremental increases.

We found no significant differencess between water based recreationists (users) and non-users (Figure 2). Moreover, our results were nearly identical to those obtained from a similar survey conducted in Montana (Kruskal-Wallis rank-sum test, $p=0.94$, Table 7.2; Suplee et al. 2009).

Discussion

This survey revealed remarkable consistency in the algae conditions that Utahns consider undesirable. We hypothesized that users would be less tolerant that non-users due to the direct association and increased familiarities, however this was not the case. The somewhat surprising agreement among these two groups suggests that these results are a broad reflection of Utah citizens. The generality of these results is also bolstered by the marked similarity between Utah and Montana surveys (Table 1) and suggests that there may be broad consensus of desirable stream conditions in the intermountain west.

Our two thresholds provide some insight into benthic algae densities that may be protective of recreational uses. The higher threshold captures a drop to baseline conditions and clearly represents degraded aesthetics. In contrast, the first threshold captures the initiation of decline in aesthetics and is therefore closer to values that are potentially protective of recreation uses. These thresholds are supported by other studies. Based on a review of several investigations, Dodds and Welch (2000) concluded that undesirable algae densities generally fall between 100 and 200 mg chl-a/m^2. Such consistency suggests that these thresholds are not merely an artifact of our study design or the specific images that we showed survey participants. However, benthic algae blooms of this magnitude almost always require both excessive nutrients and habitat that is favorable to algae growth (i.e., sufficient light, sufficient length of time between scouring floods).

As previously mentioned, Utah has traditionally protected recreation uses exclusively with *E. coli*, which only considers human health threats. These survey responses provide the information necessary to expand our protection of recreation to include the protection of aesthetics. The ability to quantitatively relate nutrient enrichment to a loss of aesthetics is important because degradation of aesthetics is more likely to alter future recreation decisions because people are generally unfamiliar with pathogen concentrations at specific waters, whereas visual problems have the potential to more directly affect everyone. Together, pathogens and aesthetics paint a more complete picture of recreation use support. These data will also help bolster Utah's nutrient reduction strategy because excessive algae growth can also cause other water quality problems, with the potential to degrade aquatic life beneficial uses.

CHAPTER 8

SUMMARY OF STRESSOR-REPONSE INDICATORS

Introduction

Utah's Division of Water Quality (UDWQ) has spent considerable effort developing multiple lines of evidence to inform the derivation of numeric nutrient criteria (NNC). The US Environmental Protection Agency (USEPA) recommends three approaches for developing NNC including: stressor-response analysis, reference condition distributional approaches and mechanistic modeling (USEPA 2010). Previous chapters describe UDWQ's efforts to use stressor-response approaches to derive thresholds for TN and TP. Several lines of evidence were explored including both functional responses (see Chapters 2-5) and structural responses (Chapter 6). In this chapter we summarize the results of these stressor-response analyses. Also, we explore, mostly as additional context to other thresholds, NNC that would be derived using the distribution of N and P among streams in reference condition.

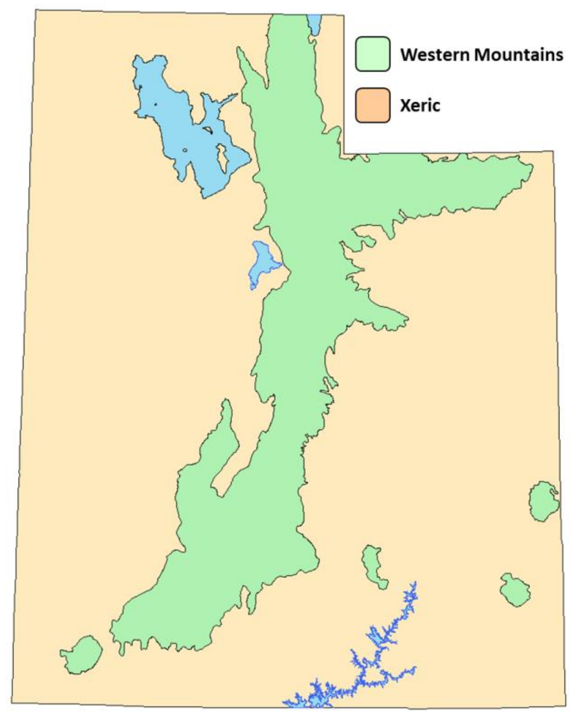


Figure 8.1. Aggregated Western Mountains (WMT) and Xeric Omernik Ecoregions used in reference site analysis.

Nutrient Thresholds from Reference Sites N and P Distributions

The USEPA recommends reference condition analysis as one approach to developing NNC. Under this approach, numerous reference sites—streams that have experienced minimum anthropogenic disturbance —are identified and sampled with the assumption that together the sites capture the diversity of streams in a predefined region. One can then develop benchmarks to delineate condition classes with a statistical evaluation of the natural variation observed among all reference sites for parameters of interest. The values used to demark classes are somewhat arbitrary, but USEPA methods prescribe benchmarks for "Fair" condition at the 75th percentile and one for "Poor" condition at the 95th percentile of all values observed among all reference sites. Implicit in this approach is the implication that some reference sites must have been misclassified prior to analysis, otherwise there is no reason for 5% of reference sites to be in "Poor" condition. Distributional approaches allow managers to easily establish benchmark over broad geographic regions, which help with generalizations of background nutrient concentrations. However, distribution approaches do not speak to the effects of increasing nutrients on designated uses, and are therefore used by UDWQ to provide context to empirically thresholds derived from stressor-response relationships. For more complete documentation of the strengths and weaknesses of this approach as well as the management implications see Herlihy et al. (2008), Paulsen et al. (2008) and Hawkins et al. (2010).

UDWQ was able to find reference data for 109 reference sites throughout Utah. These sites are continually augmented with data obtained from yearly probabilistic surveys. In addition data were obtained for an additional nine reference sites from USEPA's Wadeable Streams Assessment (WSA) program. Distributions were analyzed independently for two major aggregated Omernik ecoregions, the Western Mountains (WMT) and Xeric (Figure 8.1).

A Summary of Stressor Response Relationships

Functional Indicators

Ecosystem functions are processes that occur in waterbodies that quantify interactions between the activities of biota and the flux of matter and energy between organic and inorganic pools (Naeem 2002). Specific examples include: diel oxygen cycling, breakdown of organic matter inputs, and rates of nutrient uptake by biota. The rates and fluxes of these ecosystem processes are directly influenced by the concentrations of nutrients, so these processes quantify more immediate or direct, measures of nutrient effects than less direct structural responses (i.e., changes to fish or macroinvertebrate assemblages). Hence, functional responses can be considered an early indicator of ecological responses to cultural eutrophication. Such early indicators are important because the Clean Water Act requires that NNC are protective, meaning that they are sufficiently low to preclude degradation of uses. If used collectively functional responses can only inform NNC. However, decisions about specific N and P concentrations to use for NNC are still required because sufficiently protective concentrations are derived from consideration of what constitutes an "acceptable" loss of ecosystem structure or function, which is more a matter of both policy and science. Ultimately, any NNC that is selected will be over- or under-protective for specific sites. The specific criteria that are selected ultimately depends on the risk that resource managers are willing to accept.

We established numeric TN and TP thresholds for three measures of ecosystem function (Table 8.1 and Figure 8.2) including: nutrient limitation (Chapter 3), stream metabolism (Chapter 4) and organic matter standing stocks (Chapter 5). Nutrient limitation resulted in intermediate TN and TP concentrations in comparison to the other responses that were evaluated. Two thresholds were established for both metabolism and organic matter standing stocks, one sensitive threshold that attempts to identify early ecological responses and another that delineates where the indicator suggests highly altered conditions. Interestingly, for both responses the sensitive thresholds were among the lowest TN and TP thresholds observed across all indicators, whereas the higher thresholds were among the highest. The fact that these bracketed all indicators is useful in an assessment context because they allow UDWQ to quickly and easily identify sites that clearly have significantly altered functions (above higher thresholds) and those that clearly do not (below the lower thresholds). Sites with intermediate responses may simply require more intensive investigation.

As previously mentioned, these specific functional responses are not intended to provide comprehensive measures of functional conditions. Instead they were selected because they are relatively easy to collect and can therefore be reasonably integrated into a routine monitoring and assessment program. UDWQ anticipates that other functional responses, for instance nutrient spiraling metrics, will be developed as Utah's nutrient reduction strategy progresses. Also, UDWQ hope to augment these existing indicators to develop more robust assessment tools. Such future modifications will likely be increasingly important as UDWQ develops site-specific numeric criteria for TN and TP.

Structural Indicators

Bioassessments are one of the most common approaches that resource manages use to directly measure the health, or biological integrity, of streams, lakes and wetlands (USEPA 2002). Arguably, bioassessments are among the more meaningful assessment tools available on both sociological and ecological grounds. Members of the public often respond to measures of biological degradation more so than chemical constituents, which are inherently difficult for untrained individuals to interpret. Ecologically speaking, aquatic communities are subjected to long term impacts (weeks to years) so they can capture the effects of temporally variable pollutants. Also biota are subjected to multiple stressors (i.e., chemical pollutants, habitat perturbations) and provide an integrative measure of biological health. With respect to nutrients, in all but the most severe cases, biota are often not responding directly to stress causes by increased N or P. Instead, biota respond to increased nutrients via alterations to ecosystem functions that, in turn, negatively impact important aspects of their life history (i.e., minimum dissolved oxygen concentrations).

Several algal and macroinvertebrate biological responses, or structural indicators, were used to develop several TN and TP thresholds (Table 8.1, Figure 8.2). Data limitations precluded our ability to evaluate diatom responses to N, but the threshold for TP (0.045 mg/L), was intermediate among all biological responses. For macroinvertebrates, the most sensitive taxa started showing significant responses at very low concentrations of nutrients (average TITAN thresholds for these taxa: TN = ~0.18 mg/L, TP= 0.011 mg/L). Other more tolerant taxa, those that actually thrive under moderately high nutrient enrichment, showed significant responses at an average of about 0.6 mg/L TP and 0.41 mg/L TN. Thresholds that were directly coupled to measure of biological degradation (O/E) that have already been established by UDWQ were fairly intermediate among responses at ~0.4 mg/L TN and~ 0.04 mg/L TP, although these taxa are likely responding to other stressors that covary with nutrients.

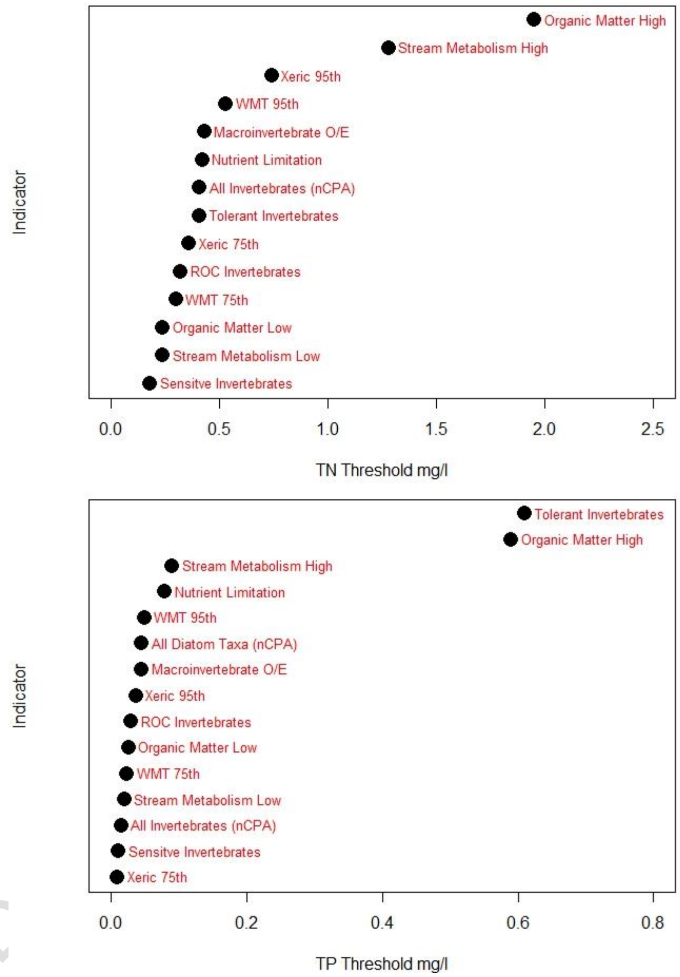


Figure 8.2. Thresholds of all nutrient indicators for TN and TP that were derived through the investigations described in this report. Indicators were developed using stressor-response analysis (functional and structural indicators) and distribution of nutrient data at reference sites (Western Mountains sites TP (n=66) TN (n=19) Xeric sites TP (n=42) TN (n=11).

While bioassessments and associated nutrient responses are useful, they have their limitations. Chief among these is the inability for these assessments to differentiate nutrient concentration problems from other sources of human-caused stress (Chapter 11). Some of the species-specific thresholds derived with TITAN have the potential in helping to distinguish stress caused by nutrients from other stressors, if similar analyses conducted for other stressors reveal different taxa-specific sensitivities. Future work on structural responses to nutrients would also benefit from a more comprehensive analysis of diatom responses because these assemblages have physiological processes that are more directly coupled to water column N and P. Similarly, macrophytes, when they are present, may help to elucidate the relative importance of nutrients stored in sediments versus within the water column. All told, however, such additional studies are unlikely to result in significantly different regional indicators. Instead, the value of these improvements will likely be manifest as site- specific standards are developed throughout Utah.

Table 8.1. List of all nutrient thresholds developed by the Utah. Thresholds were developed using stressor-response analysis (functional and structural indicators) reference site analysis (WMT sites TP (n=66) TN (n=19) Xeric sites TP (n=42) TN (n=11).

Indicators	Thresholds	
	TN (mg/l)	TP (mg/l)
Functional Indicator		
Nutrient Limitation	0.42	0.080
Stream Metabolism		
Low Threshold	0.24	0.020
High Threshold	1.28	0.090
Organic matter Storage		
Low Threshold	0.24	0.026
High Threshold	1.95	0.590
Structural Indicator		
TITAN		
Sensitve Macroinvertebrates	0.18	0.011
Tolerant Macroinvertebrates	0.41	0.610
All Macroinvertebrates (nCPA)	0.41	0.015
All Diatom Taxa (nCPA)		0.045
Biologic Assesments		
Macroinvertebrate O/E	0.43	0.045
ROC Threshold Ananlysis	0.32	0.030
Reference Site Analysis		
Western Mountains		
75th Percentile	0.30	0.024
95th Percentile	0.53	0.049
Xeric		
75th Percentile	0.36	0.010
95th Percentile	0.74	0.037

The Importance of both Structural and Functional Indicators

Among all functional and structural responses, the range of thresholds derived for TN and TP was relatively narrow. Overall TN stressor-response thresholds ranged from 0.18 to 1.95 mg/L, but the vast majority occurred at ~0.3 to 0.4 mg/L. Similar trends were observed for TP thresholds, which ranged from 0.01 to 0.6 mg/L, with the vast majority occurring at ~0.045 mg/L. This narrow range of responses will be used by UDWQ to translate these results into NNC for headwater streams.

Our multiple line of evidence approach has several strengths. First, the combined use of several measures of ecosystem structure and function provides a more comprehensive picture of the effects of increasing nutrients on elements of biological integrity. Second, as previously discussed, the deleterious effects of nutrients to stream biota can follow several different paths, and the use of multiple indicators will ultimately help to elucidate the relative importance of different nutrient responses. The latter two points highlight the principal reason that UDWQ believes that it is inappropriate for any one of these responses to be interpreted independently from the others. Finally, the more comprehensive information gleaned from multiple lines of evidence will also help to identify data gaps as we apply these approaches to the development of site-specific NNC.

The uncertainties raised by the studies presented in this report will be further addressed as detailed assessment methods are developed. UDWQ will work to summarize the results in a manner that makes them more easily conveyed to managers and the public. Finally, UDWQ will handle errors associated with transferring regional generalizations to local conditions by confirming (or deriving new) thresholds on a site-specific basis. These site-specific confirmations are particularly important in more populated watersheds where the economic implications of NNC are larger.

Next Steps: Site-Specific Investigations

The thresholds derived for nutrients and responses in the first couple of sections in this report are based on regional correlations between nutrients and ecological responses. These regional relationships are useful to agencies like UDWQ because they allow inferences at a statewide-scale, which corresponds to the scale of our regulatory responsibility. Practically speaking, it is simply not possible to manage resources at this scale while simultaneously considering specific characteristics that may make the responses of individual sites unique. However, regulatory actions are ultimately applied locally, either to specific streams or facilities. Before such actions are taken it is often important to understand the extent to which regional patterns are locally applicable. Typically these local validations are conducted in the context of allocating loads for TMDLs. However, nutrients are different because appropriate goals likely differ from place-to-place. As a result, UDWQ has opted to limit regional criteria to headwaters, which generally have lower nutrient concentrations and circumstances such that background conditions are typically attainable and appropriate water quality goals. Elsewhere, UDWQ will establish site-specific criteria for priority watersheds. These site-specific investigations, in combination with more robust monitoring and assessment of nutrients and associated ecological responses, will continue to provide insights into circumstances where regional response thresholds can be more broadly applied and circumstances where they cannot. The technical basis for the application regional stressor-response relationships to specific regulatory programs is provide in the third section of this report.

Application to Water Quality Regulatory Programs

SECTION 3

Chapter 9

OVERVIEW AND BACKGROUND

The specific application of all of the nutrient thresholds and ecological responses to Utah Division of Water Quality (UDWQ) water quality programs is of keen interest to stakeholders. Specifically, stakeholders are interested in how this information will be used to inform monitoring, assessments, numeric nutrient criteria (NNC) development, Total Maximum Daily Loads (TMDLs) and the development of permit limits. Detailed policies associated with the application of nutrient stressor-response relationships to each of these programs will be described elsewhere. Instead, this chapter provides a broad overview of potential applications of the data from the first two sections of the report. As specific nutrient reduction programs continue to be developed, this section of the report will serve as a repository for technical details and analyses that provide their technical basis.

Currently, technical considerations for program application are limited to three chapters, but we anticipate adding chapters as the nutrient reduction strategy continues to develop. Chapter 10 provides information in support of NNC development for headwater streams. This includes an exploration of the need for additional subclasses to account for natural variation in N and P, and a comparison of headwater nutrient data with statewide observations. Chapter 11 shifts the focus from the development and interpretation of regional indicators toward site-specific efforts. Specifically, this chapter describes critically important considerations for the development of robust and defensible site-specific investigations by better accounting for the influence of covariates and multiple stressors on ecological responses. Also discussed is the potential need for new responses or modifications to existing responses to better account for finer scale patterns of temporal and spatial variation that are increasingly important when assessing responses for specific sites. Chapter 12 provides the results of a study that was conducted in conjunction with the functional indicator pilot study that developed procedures for using process-based models (e.g., Qual2K) to support nutrient criteria development.

Monitoring and Assessment

Monitoring

The development of the stressor-response relationships discussed in this report involved both repurposing existing monitoring data, and the development of new monitoring and analytical methods, particularly for several functional responses (see Appendix A for data acquisition SOPs). Fortunately, the decision to augment UDWQ's monitoring programs involved careful consideration of future monitoring logistics and the need to ultimately collect data for new indicators routinely. Additionally, there are several nutrient-related responses with existing numeric criteria (i.e., DO and pH) that were not directly discussed in this report, but will continue to be independently monitored and assessed. Details plans for integrating the new monitoring elements with responses that have been historically collected will be included in UDWQ's *Strategic Monitoring Plan*. Here we describe a broad overview of how this will work.

Utah currently uses a tiered, rotating basin (six major basins) monitoring approach that combines the strengths of both systematic and random sites selection. Randomly selected sites are used for routine assessment purposed, whereas systematic monitoring is used to support regulatory programs. Each of the six major basins is visited on two different years within each rotation. On year one, 50 stream segments within the basin are selected using a stratified-random (GRTS) draw from all perennial streams watershed. At each randomly selected site, biological assessments are conducted from summer through early fall (i.e., the growing season). Monitoring at these locations includes collections of water chemistry, habitat, macroinvertebrates, diatoms, benthic algal abundance and fish. We propose adding the collection of high frequency DO and temperature for the purpose of obtaining metabolism Indicators. To meet this objective, sondes will be deployed while crews are during the initial collection effort, and then will be retrieved in batches by another staff member 7-20 days following deployment. Once a rotation is complete following implementation of high frequency data UDWQ will have a wealth of data (300 randomly sampled stream) that can be used to better understand natural variation in GPP and ER, which should provide more refined threshold responses for these indicators. All told, numerous indicators will be available at these sites that can be used to inform the nutrient reduction strategy, including: nutrients, biological assessments derived from fish, macroinvertebrates and diatoms (in development), the density of benthic algae, Gross Primary Production (GPP), Ecosystem Respiration (ER) and numerous habitat characteristics. Organic matter standing stock data are more time consuming to collect, so these data will be collected as follow-up to sites with nutrient-related problems, particularly those associated with low DO.

Data from these indicators will be evaluated and if sites are determined to threaten by cultural eutrophication, more intensive monitoring can be conducted during the "intensive" or programmatic monitoring that occurs on year three of the rotating basin schedule. In particular, this will allow UDWQ to obtain additional water chemistry samples, which would be too limited if they were based on the 1-2 samples obtained from each biological assessment site. Other follow-up monitoring for the purposed of site-specific standard or TMDL development will also be incorporated into intensive monitoring schedules.

Nutrient-Specific Assessments

In accordance with USEPA integrated report guidance, UDWQ will propose nutrient-specific assessments to be conducted in conjunction with Utah's *Integrated Report*. Once NNC are established for headwater streams, assessments will be conducted in accordance with the assessment objectives specified in the criteria and associated implementation materials. For streams lower in watersheds, nutrient assessments will be tied to Utah's narrative criteria. Assessing sites based on support of the narrative criterion is consistent with UDWQ's ongoing biological assessment program and USEPA guidance. Further support is provided by a recently adopted clarification to Utah's narrative criterion which requires that "waters of the state shall be free from human-induced stressors that will degrade the beneficial uses" and explicitly states that biological assessments can be used to determine whether uses are supported (UAC R317-7.2) We envision that quantitative nutrient assessments will be derived from collective interpretation of numeric causal and response indicators that are derived from the stressor-response relationships described in this report. We currently plan to place sites with nutrient-related impairments in a subclass of impaired waters for Utah's 303d list. As with all impaired waters, a TMDL may ultimately be required for nutrient-related impairments. However, the intent of the sub-categorization is to encourage TMDL alternatives in appropriate circumstances. Of particular importance are habitat-limited sites that are in multiple stressor environments. Under these circumstance, follow-up investigations will first emphasize the relative roles of all human stressors and, if appropriate, the best attainable conditions for the waterbody.

Specific assessment methods are under development and will be published, and open for public comment, as part of the analytical methods in Utah's *Integrated Report* when they are complete. In developing these methods, we anticipate using a risk-based approach that considers the magnitude of excursions above (or below) numeric indicators for both causal and response parameters. Assessments for parameters with existing numeric criteria, like pH and DO, will remain independent. In contrast, the functional indicators that were developed with this pilot study will be interpreted collectively. These assessment procedures will also avoid the oversimplification that results from averaging or otherwise combining responses that measure different ecological processes. Just as no single indicator among the many we developed that can conclusively determine a nutrient impairment alone, a combined score oversimplifies and eliminates many of the advantages of using multiple lines of evidence to more accurately quantify nutrient-related ecological responses. That said, the final assessment methods will also need to be sufficiently parsimonious that assessment decisions can be made consistently, following methods that can be easily communicated to stakeholders.

Development of Numeric Nutrient Criteria

From the beginning, the stressor-response pilot study and analyses were primarily intended to inform water quality criteria. The project gradually evolved to be more encompassing as Utah's nutrient reduction strategy developed, but establishing NNC remains a program priority. This section provides a general description of how we envision using the stressor-response relationships in this report to derive NNC on an iterative basis.

Numeric Criteria: Headwaters Streams

The most immediate applications of the stressor-response relationships to NNC will be the development of numeric N and P criteria for headwater streams—those classified with Category 1 or 2 Antidegradation Protections (UAR R317-2). Chapter 10 provides several technical details and analyses in support of headwater NNC development. These analyses include an evaluation of whether headwater streams require further classification and also a comparison of the distribution of headwater N and P with those predicted to occur among all of Utah's streams. These analyses, together with multiple stressor-response thresholds will provide the requisite information for the derivation of headwater NNC.

A separate document (*in preparation*), provides specific recommendations for magnitude, duration and frequency of TN and TP concentrations that UDWQ believes to be protective of the existing aquatic life uses in headwater streams. Also included is a proposal to combine nutrients with several ecological responses evaluated in this report. The decision to propose a combination of nutrients and ecological responses is predicated on the assumption that while regional criteria are broadly protective of headwater streams, circumstances likely exist where the physical characteristics of specific sites make the stream naturally resilient or sensitive to nutrient enrichment. In accordance with recent guidance (USEAP 2013) on combined criteria, this document also reviews how the combined criteria will apply to UDWQ regulatory programs.

Development of Site-Specific NNC for Streams Lower in Watersheds

The decision to establish NNC for headwater streams means that NNC will need to be developed on a continuous basis elsewhere. These ongoing investigations will require site prioritization based on a combination of the availability of data and regulatory needs. Specific approaches for site-specific criteria development will depend on factors such as: the nature and extent of biological degradation, the potential for the effects of nutrients to be confounded with those of other human-caused stressors, and regulatory needs.

While specific approaches will differ, UDWQ anticipates that investigations to support site-specific standards will generally follow procedures similar to those outlined in USEPA's CADDIS program (USEPA 2010), albeit with a modified objective. Specifically, all evidence regarding nutrients—and potentially other human caused stressors—and ecological responses will be compiled for the site. This evidence will then be organized into a conceptual model that describes all applicable pathways between nutrients and the ecological responses of interest. In planning stages, these models will primarily be used to identify pathways and responses of greatest interest and associated data gaps. The data review will then be used develop a Sample and Analysis Plan (SAPs) that defines specific approaches with respect to data management, data acquisition and proposed analytical approaches. As SAPs are developed, care should be taken to account for as many sources of potential uncertainty as possible including the spatial and temporal variation of both causes and responses, and covariates with a strong likelihood of influencing causal inferences (see Chapter 11 for a more complete review). In most cases, the development and review of SAPs and subsequent project implementations will be a collaborative process, involving UDWQ, USEPA, and other engaged stakeholders.

SITE-SPECIFIC STANDARD DEVELOPMENT: ALTERNATIVE SCENARIOS

Regionally derived stressor-response thresholds for nutrients and response parameters may need to be modified on a site-specific basis, particularly if specific responses are to be used to define water quality goals for a stream. Nevertheless, a careful review of causal and response parameters can be helpful in estimating the likely direction and scope of site-specific standard efforts. To illustrate the utility of multiple lines of evidence, this section provides several alternative approaches for site-specific investigations based on observations of both causal and response parameters. However, these scenarios are not inclusive, nor are the examples exhaustive because each alternative intrinsically involves many details that will need to be determined on a case-by-case basis.

Nutrients are High, but Ecological Responses Indicate Healthy Conditions

In some circumstances N or P may exceed regional numeric indicators, yet structural and functional responses suggest that the biological integrity of the stream is high. Initially, site-specific NNC for these streams will be established from background conditions, under the regulatory authority of antidegradation provisions, to ensure ongoing protection from cultural eutrophication. In comparison with other scenarios the level of effort under these circumstances is relatively small because once it has been established that existing conditions are fully protective, subsequent collection efforts can focus on ensuring that proposed NNC are derived from samples that represent the temporal variation of N or P at the site. NNC that are established to be protective of current conditions do not necessarily preclude future proposals to increase nutrients in the stream. However, project proponents would need to demonstrate that the increases were necessary to accommodate "important economic or social development" and the "instream water uses shall remain protective" (UAR R317-2-3.1). To demonstrate the latter the project proponent would need to provide justification that high site-specific N or P criteria would remain protective of the streams existing uses.

Nutrients are High, and Ecological Responses suggest Potential Degradation of Uses

In many ongoing monitoring data may reveal regionally-derived numeric indicators for either N or P are high and one or regional responses suggest degradation of existing uses. When this occurs, site-specific investigations should first confirm that the responses are wholly or partially the result of excess nutrients. In these circumstances, SAPs may be developed that focus on the relative role of multiple stressors in degrading responses, or on the extent to which natural conditions (covariates) may be exacerbating deleterious responses.

Once responses have been prioritized and associated thresholds confirmed or modified to accommodate naturally occurring site-specific conditions they can be used to inform NNC and ultimately, to the extent the N or P is a causal parameter, TMDL endpoints. Meeting this objective may require development of a second SAP to obtain the data necessary to couple empirical ecological responses with process-based (mechanistic) models. UDWQ and Utah State University recently completed an investigation that explored data requirements for using the combination of Qual2K models and field observations for the purpose of generating site-specific NNC (Nielson et al. 2012). The results of this investigation included recommended data quality objectives. Also, an approach for collecting the requisite data for model calibration was established. These reports, coupled with the data acquisition SOPs for the ecological responses described in this report can be used as a starting point for the development of these SOPs. In fact, all of the responses and preliminary models are already available for the receiving waters of Utah's mechanical wastewater treatment facilities, so much of the preliminary data needed to inform the development of these site-specific investigations is already available. However, the modification of collection methods for existing responses, alternative ways to summarize existing responses, or the inclusion of novel responses should all be considered to accommodate the specific data gaps identified for each site.

Ecological Responses suggest Water Quality Problems, but Conditions are Irreversible

Circumstances will also arise where attainment of regionally derived numeric indicators for causal and response parameters is not feasible due to factors such as atypically natural conditions or irreversible habitat and hydrologic modification (40 CFR 131.10(g)). Under these conditions, the stressor-response relationships described in this report can be used to inform the development of specific causal or response water quality goals that best express the best attainable condition for the stream. To determine best attainable conditions for these streams site-specific study designs will likely need to focus on the relative roles of nutrients and other stressors, mitigation efforts that can be feasibly implemented to restore all uses, and the causal and response parameters that are most likely to provide accurate and sensitive measure of stream condition. In many cases, these investigations may reveal the need to modify the designated uses of the stream in conjunction with any NNC that are proposed.

Addressing Nutrient-Related Impairments

The stressor-response relationships developed in this report will also provide useful information to help UDWQ and our stakeholders more efficiently and effectively address streams with nutrient-related problems.

Watershed Prioritization

Nutrient sources are dispersed, the science required to understand site-specific responses to nutrient enrichment is complex, and fixing nutrient related problems can be expensive. The consequence of these complexities is that UDWQ cannot address all eutrophic watersheds simultaneously. Instead, a process for watershed prioritization is required.

UDWQ, in partnership with EPA and TetraTech, has been developing several prioritization tools. For instance, results from a recently completed economic study have been geo-referenced and incorporated into spreadsheet models that can be used to provide cost:benefit information on the economic consequence of NNC implementation on an watershed-by-watershed basis. The results from these economic models have been incorporated into a Recovery Potential Screening (RPS) Tool (Norton et al. 2009), that prioritizes watershed based on the likelihood that sufficient nutrient reduction efforts—and other related restoration efforts—will restore or improve biological integrity. The RPS combines numerous GIS-based indicators to describe three indices that each measure different aspects of whether or not restorations are likely to succeed: an Ecological index, Social Index, and a Stressor Index. The social index combines several metrics that evaluates sociopolitical aspects of restorations such as the anticipated costs of remediation efforts or the commitment of local watershed groups to implement restoration efforts. The other two components of the Recovery Potential tool are more easily related to the stressor-response relationships discussed in this report. The ecological index captures the extent of impairments, for instance the departure from expected conditions, directly aligns with multiple response parameters collectively describe several causal pathways between nutrients and uses. The stressor index evaluates the number of stressors and the extent to which they can be improved to achieve water quality goals. Among other things, the proposed site-specific approaches are intended to help address complications arising in multiple stressor environments. The alignment of the RPS with the overall nutrient reduction strategy provides a way to estimate the relative cost and complexity of site-specific criteria development for different watersheds. More importantly, these tools may provide insight into places where these resources expenditures are most likely to ultimately results in improvements to water quality.

ALTERNATIVES TO TRADITIONAL TMDLS

Integral to Utah's nutrient reduction strategy is an action oriented approach for addressing nutrient related water quality problems. Traditionally, water quality impairments are addressed with TMDLs on a pollutant-by-pollutant basis. TMDLs remain an integral part of UDWQ's water quality management strategy; however, there are several drawbacks to TMDLs in addressing N and P pollution. First, TMDLs work best in situations where all sources can be clearly demarked and quantified, which will be especially difficult with nutrient pollution given the large number of both natural and human-caused nutrient sources. Second, nutrients pollution is often accompanied with other causes of degradation, in which case considering each pollutant individually may not lead to the most efficient or effective remediation practices. Similarly, in some circumstances there may be situations where it is more cost effective to remediate the effects of nutrients than to exclusively seek reduced N or P loads. For instance, if a stream is functionally impaired due to excess primary production, it may be possible to mitigate these problems by restoring riparian ecosystems to increase channel shading. Moreover, riparian restoration efforts often has other desirably outcomes (i.e., aesthetics, erosion reductions) for recreation and aquatic life uses. Most importantly, it is likely that UDWQ will identify sites that clearly have nutrient-related problems, but do not yet have NNC. Development of NNC will take time, in the interim an approach is needed that will allow know nutrient sources—particularly those that are relatively inexpensive—to be addressed on an ongoing basis.

UDWQ is proposing that TMDL alternatives are considered as potential mechanisms for more efficiently and effectively identifying nutrient-related impairments. Our proposal is to consider and alternative process to: 1) identify all potential sources of nutrients or other stressors that may be contributing to the degradation of structural or functional responses, 2) convening appropriate stakeholders, and then 3) develop and implement incremental, watershed-specific restoration efforts with the goal of restoring ecological responses. This proposed program is action oriented, because it allows for the most cost-effective restoration efforts to begin quickly, while the science necessary to establish site-specific NNC—or TMDL endpoints—is ongoing. The program is also cooperative, because it potentially allows stakeholder to come together to find local solutions to common goals. Finally, the proposed program would requires accountability because ongoing monitoring would be required to demonstrate iterative improvements, either in direct N or P reductions or in measures of functional or structural condition. Demonstration of iterative progress would be more likely because different indicators respond to restoration efforts at different spatial and temporal scales. To be clear, this proposed approach is not intended to entirely replace TMDLs requirements, but structured endpoints and load allocations may not be needed if alternative methods can demonstrate iterative progress toward meeting water quality objectives can be demonstrated.

Process-Based Models: Support of Permit Limits and Site-Specific Standards

A study collaborative study between UDWQ and Utah State University was initiated in concert with the stressor-response investigations with the dual purpose of providing tools for site-specific NNC and improving the accuracy Waste Load Analyses (WLAs) and associated permit limits. This study generated several important products that will help integrate the empirical stressor-response information with process-based models as ongoing elements of Utah's Nutrient Reduction Strategy. First, methods were developed that describe data acquisition procedures for the purpose of populating and calibrating QUAL2Kw models. Second, procedures were developed to consistently utilize these data to calibrate models with repeatedly. Third, a sensitivity analysis was conducted that identified which model parameters are most critical to the accuracy of model predictions. The results of this investigation are provided in Chapter 12 and Appendix C.

Chapter 10

AMBIENT NITROGEN AND PHOSPHORUS IN HEADWATER STREAMS

Introduction

The Importance of Headwater Streams

Headwater streams are critically important ecosystems—both ecologically and economically. Ecologically, these streams contribute to the biological integrity of all streams by providing critical hydrological connectivity among streams across large landscapes (Freeman et al. 2007). At regional scales headwater streams are critically important for the maintenance of aquatic biodiversity (Clarke at al. 2008), in part because they are physically diverse with a corresponding diverse breadth of niches (Lowe and Likens 2005). Native fish, like Utah's cutththroat trout (*Oncorhynchus clarki*), inhabit these stream year round or migrate to these streams in early spring for spawning. In an economic context, headwater streams provide many important ecosystems services. These streams, which typically represent ~2/3 of total river miles (Freeman et al. 2007), protect downstream waters through nutrient retention (Bernhardt et al. 2005), maintenance of sediment transport (Lowe and Likens 2005) and organic matters storage and processing (Muotka and Laasonen 2002). In Utah, the majority of our water falls as mountain snow, so headwater catchments are a critical part of water storage. For over three decades UDWQ has acknowledged the importance of headwater streams and afforded them antidegradation Category 1 or 2 protections (Figure 10.1, UAC R-317-2), which precludes discharges above background concentrations.

Despite existing protections, headwater streams remain threatened (Myer et al. 2007). Two important and interrelated threats to these ecosystems are habitat degradation and anthropogenic nutrient enrichment. Finlay (2011) reviewed metabolism data collected from over 200 streams and found that primary production in human-influenced headwater streams was higher than comparable reference sites, the most degraded being 600% higher. Habitat degradation can exacerbate nutrient effects. For instance, intact riparian conditions buffer the effects of nutrient enrichment directly by decreasing Photosynthetically Active Radiation (PAR) and indirectly by maintaining habitat complexity (Greenwood and Rosemond 2011). Because unaltered headwater steams are typically nutrient poor, resident biota are adapted to these conditions and are often relatively sensitive to nutrient enrichment (Miltner and Rankin 1998). Overall, incremental degradation of headwaters are more likely to have deleterious effects on these ecosystems relative to larger streams.

Protecting Utah's Headwater Streams

Utah's population is expected to double by 2050. Given the sensitivity and importance of headwater streams, it is critically important that UDWQ devise water quality management strategies that ensure ongoing protection of these ecosystems, while also allowing for anticipated growth of Utah's population and economy. Nutrients are not the only threat to Utah's headwater streams. Other stressors such as habitat degradation and hydrologic modification are also important. IN fact, one of the challenges with NNC development is that these stressors covary with nutrients along the river continuum (i.e., the magnitude of all stressors increases from upstream to downstream locales). As a result, development of numeric nutrient criteria (NNC) can potentially afford broadly protections for headwater streams. UDWQ is proposing numeric nutrient criteria for headwater—Antidegradation category 1 and 2—streams. Elsewhere, NNC will continue to be incrementally derived on a site-specific basis.

NNC define the magnitude (concentration), duration (averaging periods) and frequency (acceptable number of violations) of nitrogen (N) or phosphorus (P) that must be maintained to support existing beneficial uses. Regional NNC, such as those that UDWQ proposes for headwaters, are typically derived from thresholds obtained from two methods: empirical stressor-response (S-R) relationships and regional distributions of N and P concentrations (USEPA 2000). The underlying assumptions of these two methods are quite different. S-R methods identify statistical thresholds for N and P concentrations that demark the largest change in various ecological responses, and intrinsically assume that these statistical ecological responses also demark degraded stream conditions. S-R methods also require that sites used to conduct the analyses encompass the range of stressor conditions and include ecological responses that vary from most sensitive to degraded conditions. In contrast, NNC derived from regional N and P distributions assume that sufficient evidence exists—from hundreds of publications—that excess nutrients degrade stream ecosystems and "acceptable" levels of enrichment are management decisions. UDWQ sees both approaches as complimentary because S-R relationships can help quantify the relative risk that incremental increases of N or P pose to the support of aquatic life uses.

Accounting for Natural Variation

Addressing natural variation—in both background concentration and ecological responses—remains a central challenge of NNC development. Background nutrient concentrations vary as a result of several physical and environmental factors such as the mineral composition of soils and bedrock, soil erosion rates, organic matter inputs from watershed runoff, channel and gradient (Smith et al. 2003). In fact, national ambient stream nutrient concentrations among reference sites vary by two orders of magnitude (Lewis et al. 1999, Clark et al. 2000). Moreover, environmental gradients can buffer or exacerbate ecological responses to nutrient enrichment, which means NNC that are neither over- nor under-protective of beneficial uses also vary (Dodds and Welch 2000). Classification minimizes natural variation by systematically grouping streams with similar physical and environmental characteristics.

In 2008 there was an effort to analyze the UDWQ's monitoring data to determine if there was sufficient data to perform a stressor-response analysis between nutrients and alterations to the composition of macroinvertebrate assemblages of wadeable streams (Paul 2009). As part of this work, several different classification schemes were evaluated. One classification that was explored was an *a priori* classification scheme that compared differences in ambient nutrient concentrations and several macroinvertebrate responses among streams within Omernik level III ecoregions. Other approaches included development of empirical models that used data collected at reference sites to predict ambient nutrient classes. These analyses were unable to identify a classification scheme that minimized among group variance in nutrients or responses on a statewide basis. However, these analyses also unveiled several limitations in underlying data, which UDWQ has subsequently tried to address. In addition, these analyses were conducted statewide, whereas headwater classifications are most urgent, given UDWQ's immediate management objectives.

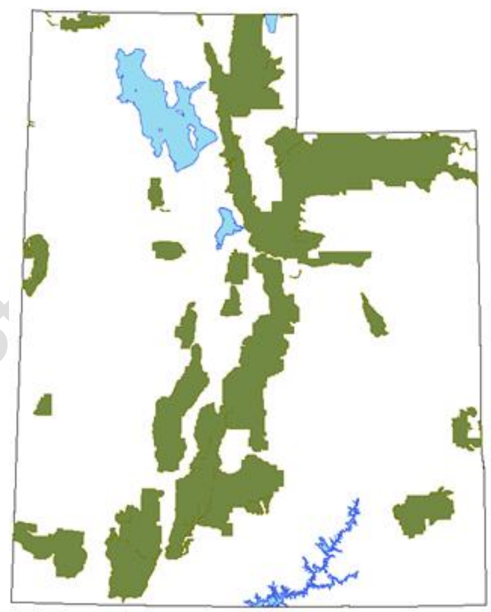


Figure10.1. Utah's Antidegradation category 1 and 2 boundaries are shown here in green. In these waters the State does not allow point source discharges (Cat 1) or only discharges equal to background concentrations (Cat 2).

Study Objectives

This chapter attempts to meet two important technical objectives of headwater NNC development. Our first objective was to determine whether or not headwater streams require further classification to account for natural differences in N or P. To meet this objective, we conduct a series of classifications and then explore between group differences in N and P ambient concentrations. Our second objective was to more broadly evaluate differences in nutrient concentrations among all headwater streams to provide context for NNC that are ultimately proposed from S-R relationships. These results, coupled with the S-R evidence discussed earlier in this report, form the underlying technical basis for headwater NNC.

Methods

Classification of all Headwater Streams

For headwater classifications, we started with relatively small watersheds as our experimental units because differentiating headwaters from streams lower in watersheds is already an *a priori* classification and we were primarily interested in whether finer-scale classifications were warranted. Hence, we established a population of all 12-digit Hydrologic Unit Code (HUC) watersheds that are currently afforded antidegradation category 1 and 2 protections. HUC-12 watersheds that overlapped category 1 and 2 antidegradation waters were also included (Figure 10.1).

Table 10.1. Watershed variables used to develop an ecologically based classification system for Utah's 12 digit HUCs that wholly and partially overlapped with antidegradtion boundaries. PCA loading are also included for each variable.resulted in 959 headwater catchments (12-digit HUCs).

			Component Loadings Model 1		Component Loadings Model 2	
Parameter Code	Source	Description	Factor 1	Factor 2	Factor 1	Factor 2
TMEAN	PRISM	Annual mean predicted from mean monthly air temperature	0.910	0.089	0.938	-0.180
LST32	PRISM	Mean day of year last freeze	-0.875	-0.032		
FST32	PRISM	Mean day of year first freeze	0.854	-0.043		
elev	Utah DEM	Mean elevation (m)	-0.788	-0.215	-0.846	-0.261
MEANP	PRISM	Annual mean predicted from mean monthly precipitation (mm)	-0.781	0.121	-0.726	0.171
XWD	PRISM	Annual mean of predicted number of days with precipitation	-0.769	0.324		
weg	STATSGO	Wind Erodibility group	-0.608	-0.176		
rockdepth	STATSGO	Depth of soil to bedrock (in)	-0.564	0.170	-0.547	0.274
awch	STATSGO	Available water capacity of soils (fraction)	0.209	0.839	0.370	-0.876
bdh	STATSGO	Soil bulk density (grams/m^3)	0.124	-0.834		
perm	STATSGO	Permeability of soild (in/hr)	0.112	-0.790	-0.083	-0.806
kfact	STATSGO	Soil erodibility factor	0.492	0.611	0.562	0.621
om	STATSGO	Organic matter content of soils (%)	-0.483	0.353		
WTAVGP	Olsen USU	Percent total phosphorus by weight of lithology	0.141	0.165		
WTAVGN	Olsen USU	Percent total nitrogen by weight of lithology	0.266	0.126		
slope	DEM	Mean slope of watershed	-0.280	0.116		
WTAVGPERM	Olsen USU	Permeability of bedrock (micrometers/sec)	0.178	-0.087		
tfact	STATSGO	Mean soil loss tolerance factor	-0.397	0.070		
HYDR	Olsen USU	Ratio of monthly mean minimum flows to mean maximum flows	0.073	0.055		
wtdepth	STATSGO	Water depth in soils (ft)	0.125	-0.054		

To develop headwater classifications, we first compiled data to describe numerous physical and environmental characteristics that quantify natural environmental gradients that are direct or indirect measures of ecologically important landscape-level characteristics. To conduct these analyses, we required data that described background gradients for all headwater streams where ambient nutrient concentrations were available. As a result, the environmental gradients that we used for these analyses are primarily based on readily-available GIS data (Table 10.1). We obtained geographic gradients of slope and elevation from Digital Elevation Models (DEMs). We also evaluated soil information soil information from the United States Geologic Survey's (USGS) STATSGO database, and lithology gradients that quantify background N and P within soils and bedrock (Olson et al. 2014). Finally, we included climate data from the PRISM database maintained by Oregon State University because background precipitation alters vegetation composition and stream hydrology. The scale of measurement for these environmental descriptors varied considerably, which can artificially under- or over-weight them in classification analyses, so each variable was normalized by subtracting the population median from each observation and then dividing by the population standard deviation prior to analyses.

Our classification analyses were iterative. We started with all possible environmental descriptors and then selected a subset of descriptors that best distinguished two groups of streams. We used k-means clustering (k=2) to compare all attributes among all watersheds. Next, we used two –factor Principle Components Analysis (PCA) to determine the relative importance of each of the 20 environmental gradient variables based on their component loading scores (contribution to each component). To control for colinearity and identify a more parsimonious solution, we identified highly correlated ($r^2>0.5$) environmental attributes and selected the variable with the highest component loading for subsequent analyses. Finally, a second K-means cluster analysis was conducted, with the resulting subset of descriptors, to identify two groups of streams that best minimized within group variability and maximized between group variations in natural environmental characteristics. PCAs were used to interpret and evaluate the relative strength of these groups.

NORMALIZATION OF N AND P DATA

We were primarily interested is whether N and P concentrations differed between these two groups of environmentally distinct streams. However, before these analyses could be conducted we needed to address a couple of artifacts created by historic laboratory methods. In the past, UDWQ analyzed samples for dissolved inorganic nitrogen (DIN, nitrate + nitrite + ammonia) instead of total nitrogen (TN). Hence, for classification purposes we used DIN instead of TN, which would have been preferable given that NNC will likely be expressed as TN. Also, until recently UDWQ's nutrient lab analysis had detection limits that were relatively high, which is problematic at reference sites where nutrient concentrations are low. To address the resulting non-detects, we used the non-parametric Kaplan-Meier (K-M) method of survival analysis (R package `NADA`, Helsel and Lee 2006). K-M method generally assumes that data are right skewed. However, water chemistry data are typically left skewed, so we used K-M "backward" to extrapolate non-detects from laboratory reporting limits to zero.

DETERMINATION OF THE NEED FOR HEADWATER SUB-CLASSIFICATION

Finally, we evaluated between group differences in DIN and TP among 99 reference sites that UDWQ had previously determined to be minimally altered by human-caused activities (Stoddard et al. 2006). Of these 99 candidate reference sites, six were dropped because they were not located in watersheds that intersected with antidegradation 1 and 2 boundaries. Another eight sites fell within the same 12-digit HUC, and we averaged the nutrient concentrations between these sites. This initial screening effort resulted in a total of 89 unique 12-digit HUCs that could use to determine whether nutrient concentrations varied among physically distinct headwater streams. Next, we coded each HUC its k-cluster group and then evaluated between -group differences in the average DIN and TP observed at each location. To determine whether N or P was statistically distinct between the classes we used a two-tailed Peto-Prentice test (NADA package), because this method is relatively unbiased by larger numbers of non-detects and the subsequent resampling that were required for these data.

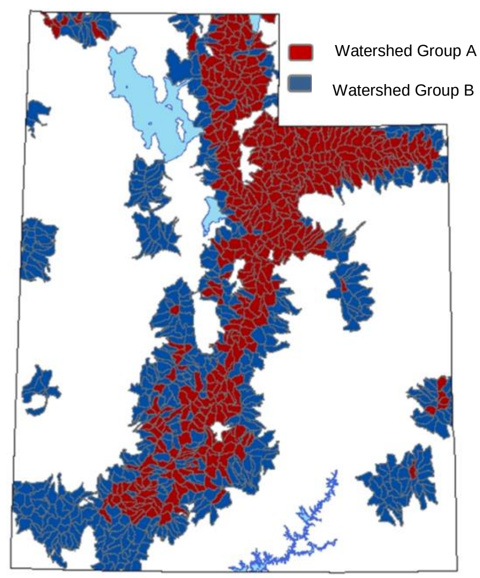


Figure 10.2. All 12 digit HUCs used in ecological classification grouped by results of k-clustering using best performing variables.

Distribution of Ambient Nutrient Concentrations among Utah's Headwaters

We expanded our classification analyses by compiling all nutrient data that were collected at all headwater (Category 1 and 2) streams by UDWQ or cooperators from 2003-2012. The total pool of stream reaches obtained from this compilation exercise varied depending on the parameter that we evaluated. One key limitation, as with the reference site classification exercise, was a paucity of TN data. However, unlike the classification exercise we thought that it was important that we express distributional results as TN, because this is how NNC will ultimately be expressed. Fortunately, we had 193 samples with both nitrate/nitrite and TN. Because these sites are in headwaters, which typically have relatively low levels of organic nitrogen relative to other nitrogen analytes. As a result, we were able to use these samples to generate a reasonably robust between relationship between nitrate-nitrite and TN ($r^2 = 0.92$, $p<0.001$). We used the linear expression from this expression to calculate predicted TN for samples where nitrate-nitrite was available, but TN was not. These predictions allowed us to examine nitrogen more broadly among headwaters. These predictions were not used for reference site classifications because we did not have data to validate that relationships held for this subset of headwater streams. Finally, we explored the distribution of headwater site average TP and TN (measured or, if necessary, predicted) among all headwater streams.

We were also interested in how ambient nutrients among headwaters streams compare with the population of all streams statewide. To make these comparisons, we used data collected from 50 streams during the summers of 2008 and 2009. These sites were randomly selected following a Generalized Random Tessellation Stratified (GRTS) survey design (Olsen and Peck 2008). This sample design differs from simple random stratified (SRS) sample designs in a couple of import respects. First, GRTS emphasizes distributing randomly selected sample locations in roughly the same way that they are distributed in the environment, whereas SRS random draws tend to be clumped. Second, the design permits stratification to ensure, in this case, that larger streams are represented despite the fact that they occur less frequently than smaller streams statewide. Most importantly, the design provides variance estimators of either stressors or responses. In this case, the GRTS design allowed us to generate, using Cumulative Distribution Functions (CDFs), spatially balanced estimates of statewide TN and TP for all streams statewide. We viewed these estimates as less biased than a compilation of all monitoring data because routine sites tend to be established for convenience (i.e., at road crossings), often to address known or suspected environmental problems, which would both tend to bias data toward higher nutrient concentrations.

Results

Classification

RELATIVE IMPORTANCE OF ENVIRONMENTAL GRADIENTS

From the initial two groups established with k-mains cluster analysis, using all 20 candidate environmental descriptors, we identified several watershed characteristics that were particularly important in describing between group differences among headwater streams (Table 10.2, Figure 10.2). These key attributes included: annual mean predicted air temperature (TMEAN), elevation (elev), annual mean predicted precipitation (MEANP), annual mean of predicted number of days with precipitation (XWD), depth of soil to bedrock (rockdepth), available water capacity of soils (awch), permeability of soils (perm) and a soil erodibility factor (kfact)(Figure 10.3). PCA axis loading revealed that the first axis, which intrinsically captures the largest portions of among site variation,

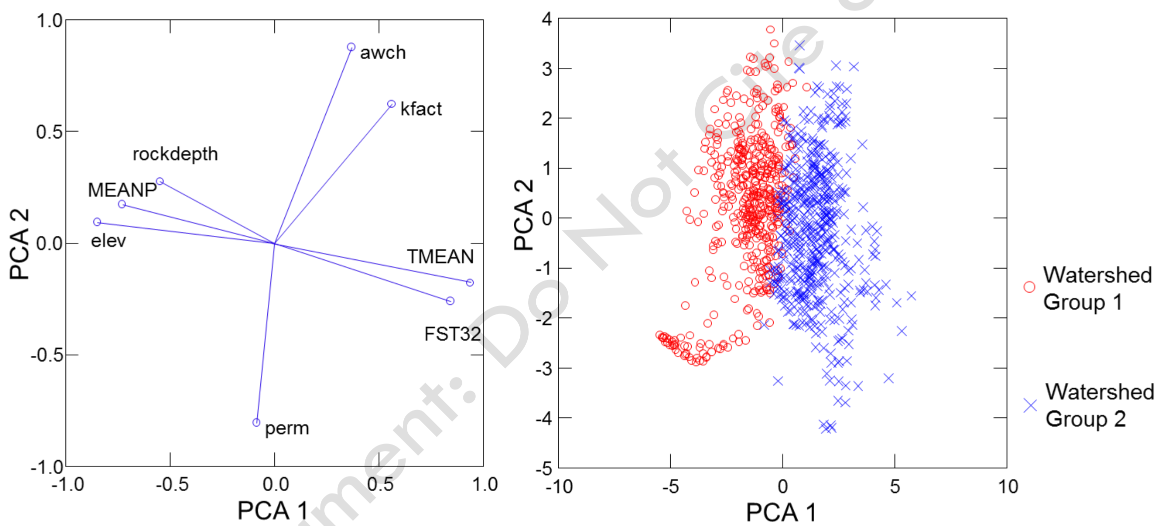


Figure 10.3. Results of PCA analysis for determining physical and environmental factors that best classify 12 digit HUCs into two distinct groups. This analysis was run with only the best performing variables from the entire population (see Table 9.1) determined by component loadings.

primarily included characteristics associated with elevation and weather (temperature and precipitation). The second PCA axis described watershed attributes associated with soil characteristics (permeability and erosion).

AMONG GROUP DIFFERENCES IN AMBIENT NUTRIENTS

Nutrient concentrations among reference sites contained a relatively large number of non-detects. About 41% of historic TIN results for these sites were below the reporting limit of 0.075 mg/L. TP chemical analyses were equally problematic with ~61% of reference samples falling below the reporting limit of 0.02 mg/L TP. All TN and TP non-detects were subsequently censored by distributing values between reporting limits and zero (R, NADA package) for purposes of these analyses.

Not surprisingly, the reference sites had fairly low nutrient concentrations. Our censored data resulted in a mean reference DIN concentration of 0.192 mg/L (95% CI = 0.125-0.259)(Table 9.2). TP was also low at these sites, with a population mean concentration of 0.017 mg/L (95% CI=0.013-0.022). Importantly, there were no significant differences between watershed groups for both DIN (p=0.906) and TP (p = 0.641) (Table 10.2).

Ambient TN and TP among Headwater Streams

The quantity of data available to evaluate the distribution of headwater ambient nutrient concentrations varied by sample location and by analyte (TN or TP). On average, each site was sampled about ~10 times for TN and ~15 times for TP over the period of record. However, the sample frequency was different among sites and ranged from one collection event (~30% of sites) to a maximum of 56 for TN and 122 events for TP. For both analytes, over half of the sites were sampled three or more times. Most of the samples were collected during the summer growing season, particularly for sites with a limited number of collection events. TP samples were available at more headwater sites (n = 605) than sites where TN was measured directly or could be estimated from nitrate-nitrite observations (n = 385). Both TP and TN (direct measure or estimates) were available for 385 headwater streams.

Table 10.2. Results of a Peto- Prentice test of significance of two watershed groups (A and B) with censored data for dissolved inorganic nitrogen and total phosphorus.

Total Phosphorus mg/l						
Watershed Group	n	n censored	Median	Mean	SD	p-value
A	46	25	0.016	0.017	0.019	0.906
B	43	29	0.007	0.016	0.018	
Dissolved Inorganic Nitrogen mg/l						
Watershed Group	n	n censored	Median	Mean	SD	p-value
A	46	17	0.099	0.237	0.424	0.641
B	43	20	0.100	0.143	0.142	

Table 10.3. Comparisons, expressed as percentiles, of headwaters and statewide ambient nutrient concentrations. Headwater distributions for TP (n = 605) and TN (n = 385) are site averages derived from all samples collected from 2002-2012. Statewide percentile estimates were obtained from cumulative distribution functions derived from samples collected at 50 randomly selected perennial streams.

	Percentiles			
	25^{th}	***50^{th}***	***75^{th}***	***90^{th}***
Total Nitrogen (mg/L)				
Headwaters	0.10	0.21	0.39	0.68
All Sites	0.18	0.25	0.50	1.1
Total Phosphorus (mg/L)				
Headwaters	BRL	0.012	0.042	0.060
All Sites	BRL	0.04	0.05	0.15

Among all headwater streams, ambient nutrients were low in comparison to statewide streams, (Table 10.3), particularly at the upper range of ambient nutrient concentrations. We found a median TP among headwater sites of 0.022 mg/L, whereas the randomly selected sites estimate the median among all Utah streams (expressed in stream miles) to be 0.04 mg/L. These differences were even more pronounced for when we compared the 90^{th} percentile of headwater streams with statewide estimates: 0.060 and 0.15 mg/L respectively. Median TN values among headwater streams (0.21 mg/L) were roughly equivalent to those expected at 50% or more stream miles (0.25 mg/L). However, headwater TN concentrations also diverged from statewide estimates at higher nutrient concentrations (90^{th} percentile, 0.68 vs. 1.1 mg/L respectively). Interestingly, sites with high TN generally did not correspond with those with high concentrations of TP.

Discussion

We stated exploring classification with GIS-based environmental gradients by identifying the two groups of headwater streams that most strongly differed in environmental attributes (Figure 10.3). Our decision to start with two groups was somewhat arbitrary, but we reasoned that finer scale classifications could be explored if between group differences in N or P concentrations were observed. We did not find differences among Utah streams, so we assumed that further delineations were unnecessary.

Correspondence with Ecoregions

Trends revealed by the PCA axis loadings were insightful. The first PCA axis revealed two important environmental gradients: 1) elevation and increasing aridity from the north to the south, and 2) those associated with soils. Given that these characteristics are important in determining background nutrient concentrations, we were somewhat surprised that these groups did not have significantly different nutrient concentrations. It is likely that those differences in nutrients among HUC-12 watersheds that do occur are the result of local attributes that are not easily captured with GIS-based data. Additional work will be needed to determine what characteristics, if any, lead to atypically high N or P within headwater watersheds. If such conditions are ultimately identified, they can be used to modify headwater NNC on a site-specific basis.

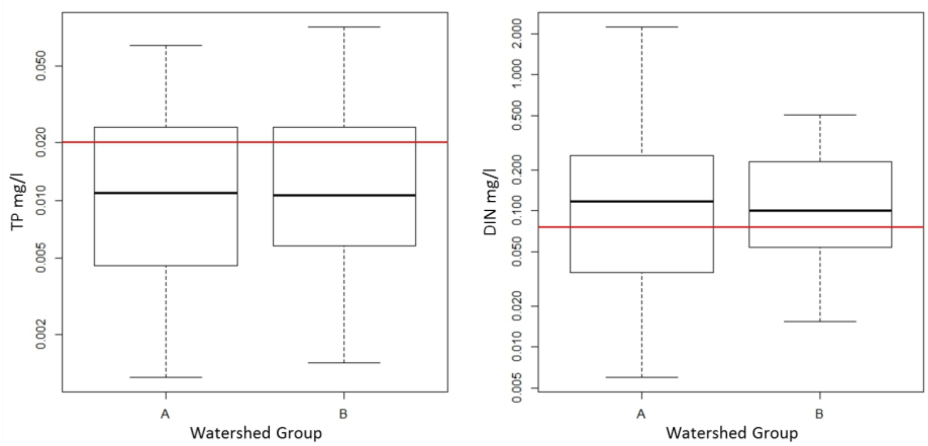


Figure 10.4. Boxplots showing distributions of total phosphorus (TP) and total inorganic nitrogen (DIN) between the two watershed groups (from k-clustering) among reference sites. Data below red vertical lines are censored and were extrapolated using Kaplan-Meier survival analysis.

Most Headwater Streams are Low in Nutrients

The variation in nutrients among reference sites was also fairly small with concentrations of TN and TP remaining consistently low. In fact, variation among headwater TP and TN (extrapolated from TIN-TN relationships) were within the 95% confident intervals of nutrient concentrations observed among headwater reference sites. Our comparison of headwaters with statewide conditions suggests that the vast majority of enriched streams occur in lower watersheds. In part, this is a natural and predictable pattern. Numerous investigators have noted that streams become increasingly autotrophic from headwaters to downstream reaches (i.e., Vanote et al. 1980). However, patterns of land-use in Utah watersheds also change predictably, with greater number of nutrient sources, often at a greater intensity, in downstream reaches. Generally speaking, covariation of natural and environmental gradients is among the central challenges of stream ecology (Allan 2000; Chapter 11), and reconciling their relative influence will remain a central challenge of Utah's nutrient reduction strategy. In fact, this challenge strongly influenced the decision to develop site-specific NNC outside of headwaters.

Overall, >90% of headwater streams fell below nutrient concentrations that are generally thought to represent a risk to stream biota (see below). This likely overestimates the extent of nutrient-related problems in Utah's headwaters because these sites were not randomly selected and in many cases were monitored to address potential water quality concerns. These results suggest that ongoing efforts to protect headwater watersheds have been mostly successful. The primary potential nutrient-related stressor in Utah streams is livestock grazing, and these data suggest that ongoing BMPs are working in the vast majority of headwaters. Once established, NNC will provide UDWQ and other agencies to better identify specific watersheds where additional management practices are needed for the ongoing maintenance of these important ecosystems.

Relevance to Headwater NNC development

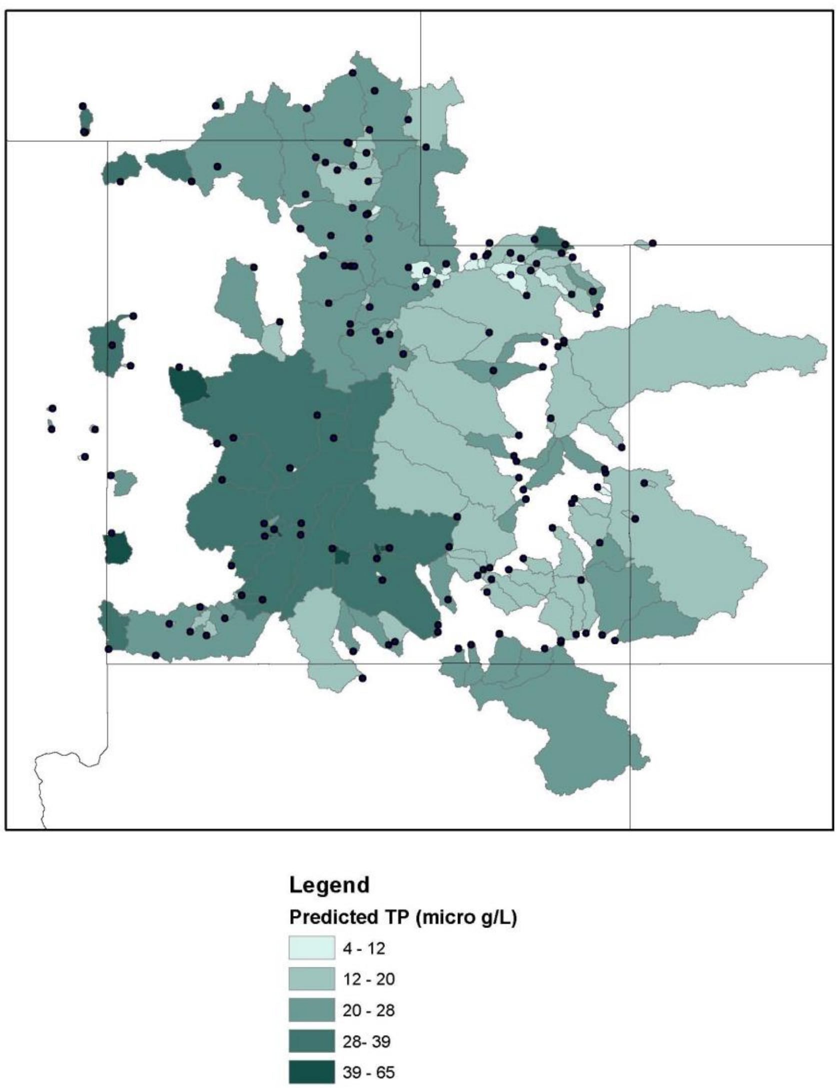
Perhaps our most important conclusion from these investigations is that headwater streams do not require further classification for purposes of NNC development. This conclusion is consistent with previous efforts to classify all streams statewide with macroinvertebrate responses (Paul 2009). This result was contrary to our expectations, particularly given climatic differences between the northern and southern regions in Utah, and its known influence on distributions of flora and fauna (Omernick 1995). These analyses suggest that, at least in the context of ambient stream nutrients, natural gradients associated with differences between mountains and valleys (e.g., temperature, gradient, size) trump those associated with climate. Local site characteristics likely exist that create locally high nutrient concentrations. However, the relatively small number of headwater streams with high nutrient concentrations suggests that such circumstances are rare. Further, at least across larger (HUC-8) scales, predictions of background nutrient concentrations fall below most S-R thresholds that we have evaluated and NNC proposed elsewhere (Figure 10.5).

These classification efforts focused on nutrients—the stressor—as opposed to ecological responses (i.e., Chapters 2-7 this report). This approach was necessary because the distribution of nutrients within headwaters was generally too narrow to account for the breadth of ecological conditions observed statewide. However, an important intrinsic assumption of this approach is that ecological responses would follow similar patterns. To the extent that nutrients cause among stream differences in responses, this assumption is valid. However, like nutrients, each of these ecological processes is known to vary naturally. We were not able to thoroughly evaluate natural variation among responses because functional responses were only available for 17 reference sites (see Section 2, this report) and of these only 7 were within Category 1 or 2 headwaters because they were intentionally selected to match the characteristics of the enriched streams in our S-R investigations. Nevertheless, none of these 7 reference streams would have been classified as "poor" condition for any of the the structural and functional responses that we evaluated.

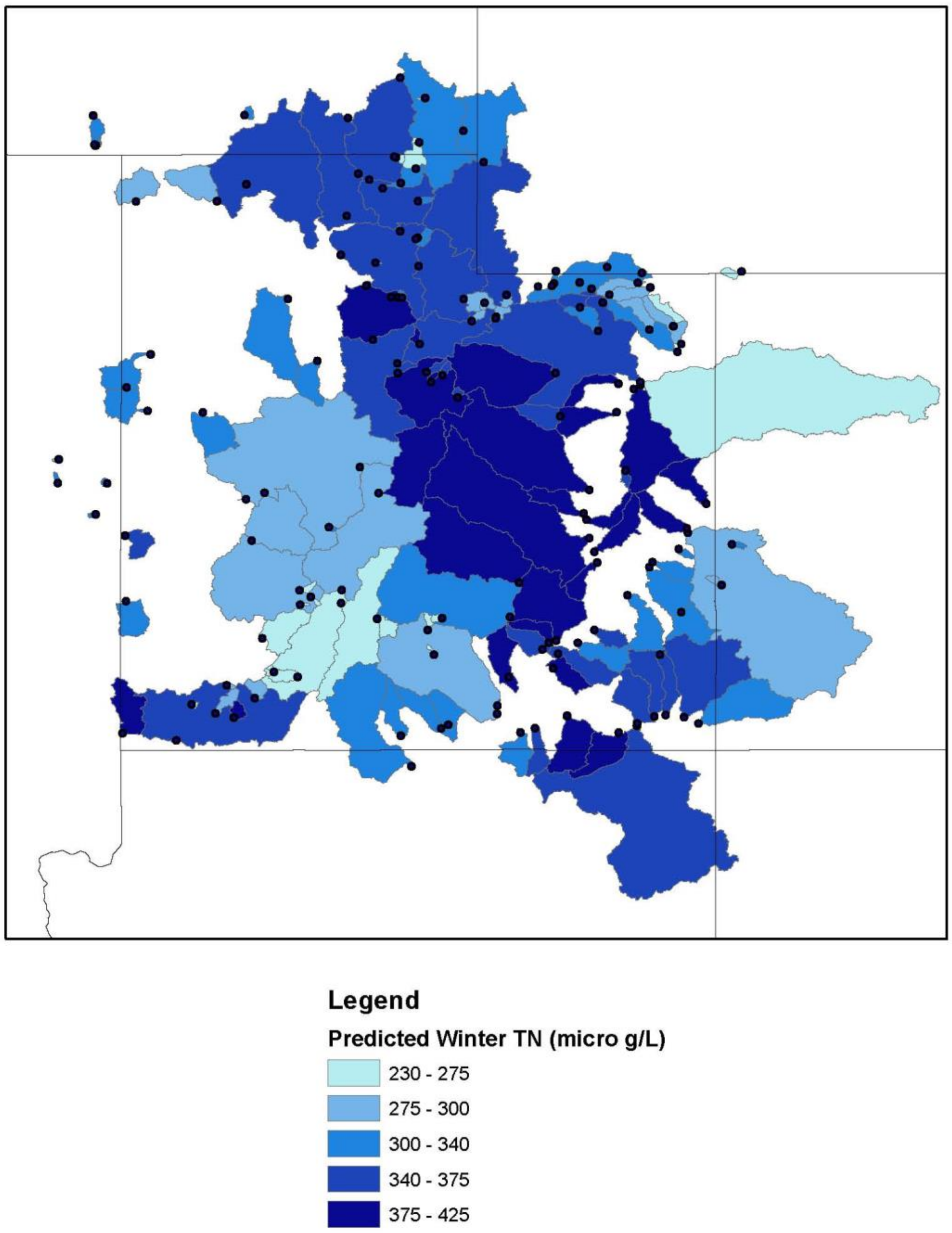
Observed reference site nutrient concentrations, and those observed among the majority of all headwater streams, generally fall within NNC, or similar protective concentrations, that others have recommended. For instance, Montana DEQ recently proposed seasonal NNC for TN at 0.25-0.325 mg/L and—with one exception of one isolated volcanic range—proposed NNC for TP ranged from 0.025-0.030 mg/L (Suplee and Watson 2013). Biggs (2000) recommended that Dissolved Inorganic Nitrogen (DIN) and Soluble Reactive Phosphorus (SRP) remain below 0.019 and 0.002 mg/L respectively to avoid nuisance algae growth (200 mg/m^2 *chl-a* for 50-day accrual). *Cladophora*, a filamentous algae that sometimes leads to nuisance algae growth in Utah streams, has a higher likelihood of reaching nuisance levels when TP exceeds 0.02-0.04 mg/L or TN exceeds 0.6-1 mg/L (Dodds 1992, Stevenson 2006), although the extent to which nuisance levels are attained depends the frequency and magnitude of floods (Freeman 1986). Collectively, these data suggest that the majority of Utah's headwater streams remain in relatively healthy condition. Future NNC for these waters will help focus resource management efforts and provide key benchmarks to ensure that they are protected for future generations.

Figure 10.5. Results of Random Forest models that predict background total nitrogen (TN) and phosphorus (TP) concentrations from observations made a reference sites under baseflow conditions (see Olson and Hawkins 2013 for details). Phosphorus predictions did not vary seasonally and predictions, displayed at HUC-8 scale, are yearly averages (Panel 1). In contrast, reference site nitrogen concentrations were consistently higher in winter (Panel 2), than summer (Panel 3), which necessitated construction of two separate models.

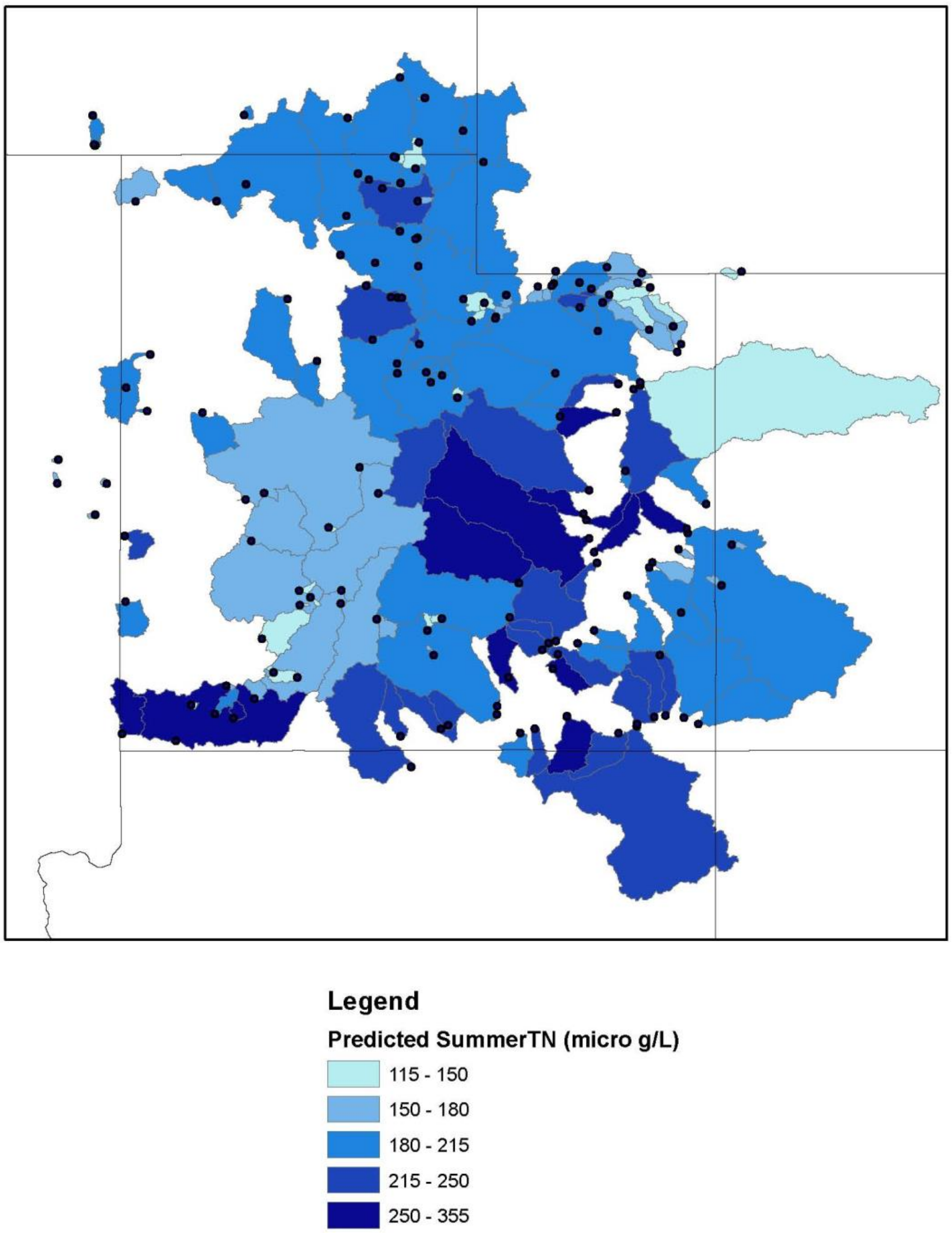
Panel A. Predicted yearly background TP concentrations.



Panel B. Predicted background TN concentrations for winter months.



Panel C. Predicted background concentrations for summer months.



Chapter 9

ACCOUNTING FOR COMPLEXITY, UNCERTAINTY, VARIABILITY, AND COVARIABLES IN SITE-SPECIFIC ANALYSES: THE PATH FORWARD

Authors: T. Miller, D. Richards and B. Marshall

Introduction

Acute and chronic thresholds for known toxicants to aquatic life were established in relatively short order and with relatively few revisions (e.g. from "Red Book" (EPA 1976) to "Gold Book" (EPA 1986)). These criteria have been a major step towards maintaining and restoring healthy aquatic ecosystems. One of the greatest remaining challenges in meeting Clean Water Act goals for 'maintaining and improving the physical, chemical, and biological integrity of our nations waters' and 'protecting and restoring the nations waters to be swimmable and fishable' is controlling anthropogenic eutrophication; which is one of the most common topics of contemporary applied aquatic ecology literature. Researchers and managers face many obstacles in establishing water quality criteria for nutrients. For example, other than rare cases of high concentrations of ammonia or very high concentrations of nitrate, many forms of nutrients are not toxic to aquatic life or human health. Rather, significant changes to important structural and functional characteristics of aquatic ecosystems can occur at much lower concentrations than are directly toxic.

One of the earliest accounts linking nutrient enrichment and algal blooms was documented by Edmondson et al. 1956 and key seminal papers were published in the 1960s (e.g. Vollenweider, 1968) and 1970s (e.g. Schindler et al. 1971). The phosphorus-phytoplankton link was well established for freshwater lakes and worldwide, P reduction programs began in the 1970s. The "P-limitation" paradigm continues today (Smith and Schindler 2009), although notable exceptions have been reported, including ephemeral or seasonal N limitations in lakes and estuaries (e.g. Howarth and

Marino 2006, Nydick, et al. 2004) and N limitation and N and P co-limitation for some streams (Dodds, et al. 2002). There are many reasons for these reported differences even within waterbody classes. These include differences in light, temperature, salinity, geology, nutrient regimes (e.g., degree of internal recycling) and resident times that seasonally favor different phytoplankton communities in lakes. Also, different algal or microbial taxa acquire and utilize N and P in different ways and early observations indicated that different environmental variables acted and interacted to produce different outcomes among different species (e.g. Mitsui et al. 1986, Tank and Dodds 2003). In addition to these factors, lotic ecosystems respond differently to varying sediment loads, propensity to scour or deposit, baseline, seasonal and spatial variability in temperature, flow and velocity, substrate particle sizes (Hynes 1970, Dewson 2007, Burdon 2013), adjacent riparian quality and stream shading (e.g. Hynes, 1970, Maloney and Weller 2011), degree and relative importance of autochthonous (within stream primary production) vs. allochthonous (external) sources of organic carbon (Dodds 2007), stream order and watershed size (Heathwaite 2010) and land use (Townsend et al. 2008, Allan 2004), among others. Many of these changes occur as natural gradients as streams flow from headwaters to downstream termini. These natural gradients of physical characteristics and concomitant biological responses results in ecosystem transitions have been defined within the context of the River Continuum Concept (Vannote et al. 1980). Human induced changes to watersheds, such as land use changes (e.g. from native forest or prairie to intensive agricultural to urbanization) can cause additional changes that have been termed human induced stressors and can include additional organic carbon and nutrient inputs, loss of pervious surfaces that cause direct delivery of sediments, various toxics, organic debris and nutrients to streams via storm drains, channelization, loss of flood plains and riparian buffer zones (Hynes 1970, Tockner et al. 2011), hydrologic modification (construction of dams and diversions ; e.g. Brooks et al. 2011). These human-induced stressors most often overlay natural transition zones and can vastly increase the variability and complexity of these altered ecosystems. In turn, this creates significant challenges for scientists and managers in sorting through these covariables to identify and prioritize causal factors of impacted (impaired) streams.

UDWQ acknowledges the that the regionally derived thresholds and indicators described in this report have limits with respect to their ability to address site-specific variation resulting from locally important covariates or with the potential for stressor-response relationships to be confounded by the influence of multiple stressor environments. As a result, Utah's nutrient strategy defines a process for establishing regional nutrient criteria and ecological responses for headwater streams, with site-specific nutrient objectives to be established elsewhere. The purpose of this chapter is to identify important sources of variability and complexity in aquatic ecosystems that interfere with empirical relationships between nutrients and ecological responses. Variability is associated with both natural stream gradients and in response to human-induced stressors. In turn, we discuss strategies to account for variability and complexity in translating regional-based numeric nutrient criteria to the development of site-specific nutrient criteria that account for natural stream transitional zones as well as multiple human-induced stressors that often accompany stream reaches that appear to be nutrient enriched. In the context of specific management objectives, we provide several considerations for the appropriate adjustments to the relative weight of ecological responses and to regional field and laboratory methods in order to account for locally important sources of natural variation. Where significant land use change and hydrologic modifications have occurred, more intensive investigation is often required to evaluate the relative importance of nutrients versus other stressors. Identifying key stressors allows the prioritization of management strategies best suited for immediate restoration efforts and can vastly increase the chances for success, both ecologically and socially.

Complexity, uncertainty, variability, and covariates

Maintaining and improving the physical, chemical, and biological (ecological) integrity of our nation's waters is extremely important and is the primary goal of the Clean Water Act. Protection and management of these waters has been delegated to management agencies such as USEPA and individual State environmental agencies (e.g., Utah Department of Environmental Quality, Division of Water Quality). Proper protection and management of waters from excess nutrients hinges on accurately assessing biological and ecological integrity and the factors that influence them; however, assessing site- specific conditions and anthropogenic effects on these ecosystems is complex and difficult, due to uncertainty, natural variability, and ever-changing ecological interactions due to natural gradients and human-induced stressors, i.e. co-variables. Aquatic ecosystems are known for their inherent complexity. All analyses of stream ecosystems are faced with elements of uncertainty and stream ecosystems are usually observed incompletely through sampling, resulting in varying levels of uncertainty (Frey 1993). Seife (2011) captures this concern in the statement that "any attempt to measure something is prone to error." The expense of sampling often prohibits collecting as much data as needed to account for and reduce uncertainty (Frey 1993). Although data may be limited to support regulatory assessments, criteria development, or site-specific analyses; decisions must be made. Therefore, the objective of resource managers is to determine the "maximum acceptable" level of uncertainty in criteria development. Such decisions require that important sources of uncertainty, particularly those associated with specific management objectives, are acknowledged and every effort made to reduce these sources to the greatest extent possible.

In addition, ecosystems are not static but are naturally variable, both spatially and temporally (Ascough et al. 2008). Virtually every biological and ecological process is variable; hence variability is a common source of uncertainty in assessments. The distinction between variability and uncertainty is not always clear and may be context dependent, particularly when there is a limited amount of empirical information available (Frey 1993; Hayes et al., 2006). Moreover, most assessments must deal with both uncertainty and variability simultaneously. Although it is tempting to ignore uncertainty and variability because they can make assessments and site-specific criteria-development challenging, they must be accounted for in the development of water quality management programs. This chapter and Appendix B aim to provide guidance to UDWQ in accounting for uncertainty when developing numeric nutrient criteria (NNC).

Addressing uncertainty and variability

In general there are three main types of uncertainty: linguistic uncertainty, epistemic uncertainty, and variability. Interactions of covariables add to the complexity and uncertainty. It is important to differentiate these types of uncertainty before they can be integrated into the various approaches to account for uncertainties during development of site-specific criteria (Ferson, 1996; Ferson and Ginzburg, 1996; Regan et al., 2002a, others).

Linguistic uncertainty comes from the difficulty in communicating what we precisely mean. Vagueness, ambiguity and context-dependence are sources of linguistic uncertainty (Gregory et al. 2012). Context is everything. Assessments and site-specific analysis can reduce linguistic uncertainty by clearly stating and defining important concepts and terms during their development (Gregory et al. 2012; Regan et al., 2002).

Epistemic uncertainty is uncertainty associated with knowledge or lack of knowledge of the state of an ecosystem (Hayes et al., 2006; Morgan and Henrion, 1990) and pervades all of our attempts to discover the truth about ecosystems and our ability to make sound assessments or to develop meaningful management criteria (Ascough et al. 2008). Uncertainty can be reduced by additional research, i.e., the parameter value can be refined and then further quantified (Frey 1993; Hayes et al., 2006). For example, refining site -specific estimates of phosphorus levels that result in 'blooms' of two nuisance algal taxa, *Cladophora* sp. and *Didymosphenia* sp. can be further quantified by additional research. There are many types of epistemic uncertainty including: measurement error; data uncertainty; systematic error; statistical uncertainty; model uncertainty; parameter uncertainty; inherent randomness; and subjective judgment (Frey 1993; Hayes et al., 2006). Data uncertainty from measurement error can occur throughout the assessment process. Systematic error is the error that is constant in repetitions of the same experiment, observation, or sampling protocol. An example of systematic error is the continued use of a functional feeding group (FFG) assignment of a species that was based on family level FFG classification or even misclassification, as opposed to a species or genus level assignment or reclassification based on additional literature review. The opposite of systematic error is random error. Systematic errors are usually much more serious than random errors because their magnitude cannot be reduced by simple repetition and these errors often go unrecognized. Additive systematic error is known as 'bias'. The only way to deal with systematic error is to recognize a bias and remove it by thorough examination and validation of the assessment procedure.

SAMPLING DESIGN

> **"No amount of statistical sophistication can rescue a poorly designed study"**
> **(Lovell 2013)**

The first step toward reducing error and uncertainty is to articulate a clear research question and accept that spatial and temporal environmental variation exists. The next step is to develop an appropriate sample design that is relevant to the research question and accounts for inherent spatial and temporal variation. Appropriate sampling design and data collection are necessary prerequisites for developing robust site-specific criteria regardless of how the data are analyzed. Stratification, randomization, and replication (SRR) are imperative in all site-specific sample designs because they reduce error and uncertainty (see The Path Forward and Appendix B for detailed description of improved macroinvertebrate sampling design). Power analyses can help determine the ideal number of samples per stratum to balance sensitivity and cost of the assessment. Although randomization is important, randomization alone is inadequate. A completely randomized design without stratification infers either that the study subject occurs completely randomly in the study area or that the researcher has no *a priori* knowledge of the subject. Nothing in nature is random and we know quite a bit about our study subjects (i.e., ecologists have actually amassed much information about the spatial and temporal distributions of organisms in streams), therefore, stratification is imperative. Stratification can also help focus assessments on specific stressors by reducing variability associated with generalized random sampling designs. Sampling design strategies for site-specific assessments that could be considered can be found in Thompson and Seber (1996) and Chao and Thompson (2001). Site specific sampling methods must also compliment the sample design and will likely differ from those used for region wide sampling methods. In addition, knowledge and avoidance of pseudo-replication

(Hurlbert 1984) is required in all site-specific sampling designs. In Appendix B, we discuss specific field and laboratory sampling methods that can improve site-specific assessments focusing on macroinvertebrates.

Given that many ecological processes evolve dynamically, both spatially and temporally, purely spatial sample designs are often not as efficient as those that consider spatiotemporal dependence (see Underwood 1996). Many methods and techniques are available that model both, simultaneously including hierarchical models such as; network designs (Le and Zidek 1994), optimal designs for time-dependent responses (Federov and Nachtsheim 1995), updating sample design in repeated environmental surveys (Arbia and Lafratta 1997), optimal network designs for spatial prediction covariance parameter estimation, and empirical prediction (Zimmerman 2006), and dynamic design networks for space-time models and non-Gaussian data (Wikle and Royle 1999, 2005).

STATISTICAL APPROACHES

Frequentist statistics are the most well-established and commonly used methodologies of statistical hypothesis testing (i.e., Null Hypothesis Significance Testing (NHST)). However NHST has come under great criticism of late (Johnson 1999, Siefe 2010, Ziliak and McClosky 2009, Goodman 2008). An alternative to Frequentist statistics is Bayesean statistics. Both schools of thought have their place and usefulness in ecological assessments and a thorough understanding of their nuances and applicability are needed. For example, when Frequentist NHST methods are considered for use, an understanding of their limitations and the appropriate use of Type I error (false positives), Type II Error (false negative) and the meaning of p-values are necessary. Statistical methods that don't rely on NHST and p –values that could be considered and evaluated for site- specific criteria include: estimation approaches based on confidence limits, maximum likelihood, quantile regression; some multivariate methods such as hierarchical clustering, non-metric multidimensional scaling, Random Forests, etc.; AIC, structured equation models (SEM), non-parametric multiplicative regression, or fundamentally different approaches based on Bayesian statistics. The selection of the most appropriate method (s) will depend on the specific question to be addresses, the level of uncertainty that is acceptable during NNC development and the parameters and data chosen for evaluation. Several of these multimetric methods are mostly exploratory in nature such as hierarchical clustering and NMDS, whereas Random Forest models, SEM, and NPMR are more confirmatory. Bayesian methods can be both exploratory and confirmatory but allow *a priori* estimates and information to be included in the analyses. Multiple lines of evidence are also quite useful and can add strong decision support, particularly if they measure several different ecological responses. As previously mentioned, the chemical, physical and biological responses to excess nutrients are varied and complex. Multiple lines of evidence that describe different aspects of these responses can be pieced together to paint a more complete picture of nutrient-related problems or other covariables that are causing the observed degradation. Such knowledge is critically important to minimizing assessment errors. More importantly, a more complete understanding of the problem frequently informs the selection of restoration activities that are most likely to efficiently and effectively achieve water quality objectives.

ENSURING BIOLOGICAL SIGNIFICANCE

Statistical significance and biological significance are not always the same. Statistical significance is almost meaningless in site-specific assessments if it is not biologically relevant. Defining what is biologically significant/relevant is not straightforward, nor is it a statistical decision. Defining biological significance has important consequences for the design, statistical analysis, and interpretation of an experiment and development of site-specific criteria (Martinez-Abrain 2008, Lovell 2013). Biologically significant effects can be quantified by determining 'effect size' and using power and sample size calculations (Martinez-Abrain 2008, Lovell 2013). The choice of an appropriate effect size however, requires a thorough understanding of the biological subject and the context in which it is being evaluated (see Appendix B). Expert judgment and knowledge by the 'domain' scientists is required and there may be no consensus as to what the minimum difference (effect size) to be considered significant is (Lovell 2013, McBride and Burgman 2012). The choice of the effect size is, therefore, a decision for the domain scientist (s) based on expert knowledge or in the case of site-specific, criteria development; the managers responsible for such development (Lovell 2013).

Power analysis is a useful tool for informing biological relevance (i.e. level of difference, effect size) because it captures variation in the context of effect size. Often the magnitude of ecological response is a more appropriate decision tool than NHST p-values, which only identify whether or not measures of central tendency differ among populations. In the case of power analyses, p-vales can be replaced by confidence intervals. Since effect size is decided by the 'domain' scientist (s), the choice of upper and lower CI bounds or CI intervals (e.g. a choice of 90% CIs because they are somewhat equivalent to $p = 0.05$) is not a statistical decision (Lovell 2013).

MODEL UNCERTAINTY

> **"Remember that all models are wrong; the practical question is how wrong they have to be to not be useful."**

Box and Draper (1987)

Models are simplified representations of a real-world system and therefore likely to be incomplete or incorrect and are often a key source of uncertainty (Frey 1993; Hayes et al., 2006). Competing models are often available and the choice and use of the most appropriate model(s) can reduce uncertainty considerably. However, each model comes with its own suite of input parameters and decision rules. For example most multivariate methods provide the researcher with a plethora of statistical model choices (e.g. type of distance matrix, maximum likelihood vs. bootstrapping, etc.). Assumptions need to be made throughout the modeling process. Many of these choices may be equally valid or may not have been thoroughly evaluated for use in ecological data. Their selection can significantly affect the outcome of the analysis. In addition, the results of these choices are multiplicative and can result in numerous combinations of models, each choice of which can result in differing model outcomes. Uncertainty imposed by choice of models and the cumulative effects of input parameters and decisions rules needs to be careful examined prior to analyses and this uncertainty needs to be addressed.

Increased model complexity does not always reduce the amount of uncertainty (Frey 1993; Hayes et al., 2006). An increased number of parameters in a model may in fact increase the uncertainty of the model outcome for a given set of data and can be either additive or multiplicative or both (McCune 2011). When there is more than one uncertain quantity, uncertainties may be statistically or functionally dependent or correlated (Cressie et al., 2009). Failure to properly model the dependence between the quantities can lead to uncertainty in the result, primarily in the variance of output variables.

The most useful models will provide the greatest simplifications while providing an adequately accurate representation of the processes affecting the phenomena of interest (Walker et al., 2003). Merow et al. (2014) suggested that researchers should constrain the complexity of their models based on study objective, attributes of the data, and an understanding of how these interact with the underlying biological processes. The purpose of a quantitative uncertainty analysis is to quantify the degree of confidence in an analysis or assessment using the most appropriate data and models available (Ferson and Ginzburg 1996, Ferson et al. 1999). Sensitivity analysis can be used to determine where uncertainty reduction is most necessary and beneficial. The use of probability and interval analysis can deal with all epistemic uncertainty (Ferson and Ginzburg 1996, Ferson et al. 1999), however the increased accuracy of probability distributions and interval analysis can result in decreased precision (Richards 2009). Accuracy is always more important than precision (Ziliak and McCloskey 2009) and if a qualitative assessment is being conducted; precision may become unnecessary or non-essential (Dambacher et al., 2003).

TECHNICAL UNCERTAINTY

In addition to the types of data uncertainty discussed, there can also be technical uncertainty. Technical uncertainty is the uncertainty generated by software or hardware errors, i.e., hidden flaws in the technical equipment (Frey 1993; Hayes et al., 2006). Software errors arise from bugs in software, design errors in algorithms, and typing errors in model source code (Walker et al. 2003). There are many new, exciting, and potentially useful statistical methods and ecological models published almost daily in well known, peer- reviewed journals or that have been developed for regulatory agencies or commercial use (e.g. every issue of Ecological Applications, Biometrics, Environmetrics, Environmental and Ecological Statistics, etc.) These methods should be used with caution for nutrient criteria development while they are in their infancy and until they have been further tested and verified.

INHERENT RANDOMNESS

Some quantities may be irreducibly random, however, this concept is often applied to quantities that can be measured precisely, but as a practical matter are not. In ecology, true randomness is rare, and inherent randomness is unlikely. Some processes that resemble randomness are actually the product of unmeasured variables or covariables; for example the distribution of macroinvertebrate taxa within riffle habitat. Several strategies that can help reduce apparent randomness are discussed later in this chapter.

MODEL OUTPUT

This is the accumulated uncertainty that is propagated through the model. Model output uncertainty is the result of all of the types of uncertainties listed above and variability, which will be discussed below. Model output uncertainty is often ignored or misunderstood. One often overlooked outcome of accumulated uncertainty occurs when the total model error rate is larger than the predetermined significance level of the model results. Model results are likely invalid when this occurs (Seife 2010). Model output error is sometimes called prediction error. Uncertainty occurs at every stage and can accumulate throughout the analyses process. Uncertainty should always be accounted for by thoroughly examining <u>every</u> stage of the analyses and reducing it when possible and by reporting results with bounds such as probabilities and confidence intervals; only then can we be confident in the scientific inferences made from an analysis (see Cressie et al., 2009).

Variability

The true nature of rivers and streams: "Patchy in space, dynamic in time"
Townsend and Hildrew 2014

Variability is a type of uncertainty that is also referred to as external, objective, random, stochastic, or natural variability (see Appendix B). It is related to the inherent variability in natural and human altered ecosystems. Understanding natural variability is critical in management decisions since it is usually poorly understood and often confused with knowledge uncertainty (epistemic uncertainty; Frey 1993; Hayes et al., 2006). Variance is a measure of the heterogeneity of and ecosystem parameter. Variance cannot be reduced by further research but it can often be represented more accurately and communicated better with additional data. Variance is best often quantified as a frequency rather than a probability distribution (Frey 1993; Hayes et al., 2006). Additional mathematical techniques that can be used to address model uncertainty and variability include: Qualitative Modeling (Dambacher et al., 2002), Bayesian Belief Networks (Henrion et al., 2001), Aikake Information Criteria (AIC), second-order Monte Carlo simulation (Cullen and Frey, 1999), Probability Bounds Analysis (Ferson, 2002), Information-Gap Theory (Regan et al., 2005) and hierarchical Bayesian techniques (Link et al., 2002).

Incorporating Covariates into Site-Specific NNC Development

In the context of this chapter, covariates include the sum of additional physical, chemical or biological attributes that may operate as natural gradients or as human-induced stressors that confound interpretations of the extent to which the stressor of interest (i.e. nutrients) causes changes in ecological responses. Many of these characteristics have been described in the context of natural gradients of variability as streams flow from headwaters to larger streams with greater flow, warmer temperature, lower gradient, etc. One of the earliest treatises on this subject was Hynes (1970). He recognized the unique tolerances of specific periphyton and macroinvertebrate taxa to natural physical gradients and variables, including those that tolerated wide ranges of variability (tolerant taxa) and those that tolerated narrower ranges (intolerant taxa). Hynes (1970) characterized these natural gradients in physical and chemical characteristics as natural and predictable zones, which favor specific taxa.

Shortly thereafter, Vannote et al. (1980) recognized that the transition between such zones occurs on a continual gradient and proposed the River Continuum Concept (RCC) as a more holistic view of watersheds where the gradient elicits a series of predictable responses within biological communities as populations follow a similar continuum of abiotic factors and energy sources and the transport, utilization and storage of organic matter and nutrients along the length of a river (i.e. nutrient spiraling (Webster et al. 1975). Within the RCC framework, streams will generally transition from more oligotrophic allochthonous primary production in headwaters to a mesotrophic state that includes increasing contribution and importance of autochthonous organic matter and nutrient loading from more developed riparian communities and larger watersheds (Vannote et al. 1980, Dodds 2007). This complexity is further complicated by seasonal transitions from primarily heterotrophy to autochthonous primary production (spring and summer) because of increased light availability and a reduction in allochthonous detritus abundance, for example, after leaf fall (Dodds 2007). As streams increase in size, there is usually a transition to lower physical gradients and smaller inorganic substrata size. In turn, this increases the ability of streams to retain, process, and store organic matter and nutrients within depositional sediments. In response, biotic communities also transition to include taxa that can occupy differing/smaller substrates to capitalize on energy (organic carbon) that originated in upstream communities. This often leads to a dominance of heterotrophic oriented taxa. These factors concurrently and continually interact to influence the habitat availability for specific taxa and the community assemblages of primary and secondary producers and macroinvertebrate populations that respond to these physical and biological patterns.

The importance of systematic changes along natural environmental gradient, like those described by the RCC, was acknowledged by EPA in their Nutrient Criteria Guidance: “A directly prescriptive approach to nutrient criteria development is not appropriate due to regional differences that exist and the lack of a clear technical understanding of the relationship between nutrients, algal growth, and other factors (e.g., flow, light, substrata)”. Importantly, however, this statement only acknowledges “regional differences” in important variables, while the examples provided (e.g., flow, light, substrata) certainly vary on a local site-specific basis. Further, in its water quality standards regulations, “EPA recommends that States and Tribes establish numerical criteria based on section 304(a) guidance, section 304(a) guidance modified to reflect site-specific conditions, or other scientifically defensible methods”.

Still, a much greater challenge lies in the fact that human activities have altered each one of these natural gradients by: (1) dewatering, (2) channelizing (3) altering substrate composition, (4) altering riparian vegetation, or (5) altering natural landscapes in the watershed. The cumulative result has been to exceed the typical ranges of normal variability of virtually all abiotic factors discussed above. The degree of alteration among these factors may be additive or even have multiplicative effects on the structure and function of biotic communities (Folt et al. 1999, Underwood, 1989).

Secondary influences can also arise from the interactions of potentially new combinations of species in anthropogenically-altered states that may result in new ecosystems or simplified regimes (Hobbs et al. 2006) that have yet to be understood and described. Indeed, many evolutionary ecologists have suggested that we have now entered a new era, known as the 'Homogocene' or the 'new Pangea' where biotic assemblages and communities within an ecosystem will no longer be discernable anywhere on the globe because of the loss of biodiversity and the homogenization of the remaining native species with introduced species and eventually a world where only pests and weeds survive (Ulansey 2014). In addition, scientists are also calling the time period in which we live the "Anthropocene", the time in which human activity has become a driving force in nature, affecting everything from the carbon cycle to the climate (Ulansey 2014). As Meyers (2004) suggests, "The extinction crisis is over. We lost." Meaningful nutrient criteria will have to address this reality by working within an adaptive management framework and future criteria modifications are likely.

The most common approach in understanding community changes of biological communities has been to establish a suite of reference sites that represent a regional average of geographical and meteorological conditions and an abbreviated taxa list that most commonly occurs among these reference sites. Subsequent sample collection of the local taxa (e.g. macroinvertebrates) from a site that falls within that ecoregion is then compared as a fraction of the species that occurred in the regional reference list. One key problem with using regional indicators in establishing reference condition is the lack of resolution and understanding of conditions that occur within major transitional zones of the river continuum. This ignores natural gradients in stream abiotic and biotic condition, which naturally dictate species distributions. For example, few Plecoptera in the Northern and Middle Rocky Mountain Ecoregions occur where summer water-temperatures exceed 20^{0}C (reference), or where sediment embeddedness exceeds about 40% or where sand or silt particles dominate the substratum (Richards unpublished data). The absence of these taxa in mid or low-elevation streams may not be due to nutrient gradients but rather to constraints by the physical conditions to which they are adapted. It would therefore be prudent that reference condition also be refined with greater resolution by identifying and stratifying subsets of regional reference sites that represent the appropriate zones according to river continuum principles that occur within that region. Ideally, these zones would incorporate an understanding of energy sources and flow within these zones. Accordingly, Dodds (2007) discussed the importance of trophic state, as incorporated within zones, which are primarily heterotrophic or autotrophic in influencing nutrient utilization and therefore different thresholds of impacts. Even within such zones, site-specific conditions greatly influence stream metabolism and algal growth (Dodds 2007, Hill et al. 2009), and other characteristics. For example, flood frequency also influences periphyton growth and accumulation and all else being equal, eutrophication effects will likely be stronger under stable flow regimes (i.e., intermediate disturbance hypothesis) (Biggs, 2000).

The Path Forward

Given the complexity of ecological systems, all sources of uncertainty in site-specific numeric nutrient criteria development cannot be removed; therefore resource managers need to decide on the degree of uncertainty that is acceptable in the context of specific management objectives. It is important to acknowledge that uncertainty that is not addressed frequently leads to conservative or even erroneous regulatory actions. As a result, it is in the interest of both UDWQ and the regulated community to ensure that uncertainty is reduced to the greatest extent possible.

There are numerous considerations in making determinations about the size and scope of site-specific investigations. An important first step in making these determinations is a careful examination of all water quality data to determine the strength and weaknesses of existing data. As an example, if the application of site-specific nutrient criteria would result in expensive wastewater treatment facility upgrades or Best Management Practice implementations, then additional monitoring and research might be warranted to reduce the risk of overly stringent criteria and associated implementation costs to local communities. One the other hand, the possibility also exists for extended deleterious impacts to streams while investigations are ongoing. Similarly, delayed regulatory decisions may result in additional legal vulnerability to UDWQ or the regulated community. Taken together, pros and cons of these decisions emphasizes the importance of the perspectives of different stakeholders and the need for trust, collaboration and consensus in reaching management objectives in both criteria development and implementation.

Further reduction in risk and uncertainty can be accomplished within the framework of adaptive management. Adaptive management is embraced by UDWQ because it intrinsically acknowledges uncertainty by continued periodic monitoring that can provide for refinement of management decisions. In fact, UDWQ's approach toward the development of numeric nutrient criteria is an example of an adaptive decision that is intended to facilitate adoption of regional numeric nutrient criteria for headwater streams, while more nuanced issues of covariates and multiple stressor environments continue to be addressed on a site-specific basis elsewhere. The decision to implement a statewide technology-based effluent limit for phosphorus of 1 mg/L is another important example of implementing an adaptive management strategy. In this case, UDWQ and stakeholders acknowledged that most wastewater treatment facilities and industrial discharges add sufficient phosphorus (P) as to pose a threat to the integrity of receiving waters. UDWQ and stakeholders agreed that the necessary treatment facility modifications were not economically prohibitive. These modifications will provide for an initial and substantial (approximately 2/3) reduction in effluent P concentrations and are expected to improve biological conditions of many receiving waters. This will nearly halt additional P loading from point sources in the near future while additional studies can be performed on priority streams to identify where additional nutrient reduction is necessary or other stressor(s) need to be mitigated in order to maintain or restore biological integrity of receiving waters. The following section incorporates topics highlighted earlier into site-specific investigations as employed through an adaptive management framework.

Development of Study Designs

Once existing data available for a site has been compiled and placed into a conceptual framework, the next step is designing a study to meet management objectives. As previously highlighted in this chapter, this step is the most important if the study results are ultimately going to account for complexity, and minimize uncertainty and variability. Here we discuss general considerations for appropriate site-specific designs, followed by specific examples of how these concepts might be applied to two important, and interrelated, questions important to the nutrient reduction strategy: the site-specific modification of nutrient indicators (responses) and the development of site-specific numeric criteria.

GENERAL CONSIDERATIONS

Many of UDWQ's routine monitoring and assessment procedures were developed to provide insight into regional conditions and water quality concerns. As site-specific designs are established, there are several contexts where routine monitoring methods should be reconsidered to better match the specific study objectives at smaller spatial scales. Modifications that warrant careful consideration of several critical study design elements including: site-selection, sampling design, field sampling methods, laboratory methods and augmenting existing indicators with metrics that are biologically relevant to nutrients at a local scale.

THE IMPORTANCE OF REPLICATION

Elements of replication are, of course, required to account for within-site variation and should be applied to all site measurements whenever possible. One way to determine the amount of replication required is power analysis, which is discussed with respect to macroinvertebrate collections in Appendix B. Wherever possible, replication should be done to account for each of the covariables addressed through a stratified sample design, which might involve both regional and site-specific characteristics.

SITE SELECTION

Most study designs will require comparing data collected at the site of immediate interest against data collected at other sites. If this is an element of the study design care must be taken to ensure that these sites inform, rather than confound, management objectives. The accumulated literature suggests that regardless of the purpose all sites should be selected to account for natural gradients that occur within single watersheds. Of particular importance are natural gradients in temperature, substrata, nutrients, allochthonous and autochthonous sources of organic carbon, and the extent to which natural changes in biological composition might be exerting top-down controls on stream biota.

In some circumstances, indicators will need to bench marked against reference sites to derive estimates of background conditions. When this is necessary care should be taken to ensure that sites are selected to account for important covariables that are known to influence ecological responses. In Utah, most second or third order streams have been fundamentally altered after flowing from the relatively protected headwaters of USDA Forest Service lands; flowing through altered landscapes (agricultural or urban), and experiencing dewatering for irrigation, and/or degraded riparian communities. Therefore, reference sites need to be identified that have minimal human disturbances, but otherwise have comparable physical and chemical characteristics as the site that is under investigation. Selection of appropriate sites will likely involve consideration of finer-scale attributes than UDWQ typically uses for regional assessments because regionally important covariables may not be the primary determinate of biological responses at finer, local scales. This is particularly true for major transitional zones along the river continuum. For example, few Plecoptera in the Northern and Middle Rocky Mountain Ecoregions occur where summer water-temperatures exceed $20^{0}C$ (reference), or where sediment embeddedness exceeds about 40% and small gravel, sand or silt particles dominate the substratum (Richards unpublished data). The absence of these taxa in mid or low-elevation streams may not be due to nutrient gradients but rather to constraints by the physical conditions to which they are adapted. It would therefore be prudent that reference condition also be refined with greater resolution by identifying and stratifying subsets of regional reference sites that represent the appropriate zones according to river continuum principles that occur within that region. This is necessary because other aspects of landscape changes may have equal or greater effects on invertebrates and periphyton assemblages than nutrients alone (e.g. Burdon et al. 2013, Wagenhoff et al. 2012, Maloney and Weller. 2011, etc.). However, identifying the mechanisms behind these influences is complicated by the many potential pathways (often indirect) between land use and ecosystems and by the long lasting effects of past land use. We need to better understand these indirect and lasting effects in order to support ecosystem restoration and conservation efforts (Mallony and Weller 2011).

Although the use of reference sites is preferred, control sites may be more appropriate and may be better able to distinguish the relative role of multiple stressors on ecological responses. Downes (2010) suggested avoiding the idea of reference sites altogether and to use control-based site evaluations. Downes (2010) and Quinn and Keough (2002) suggested that the purpose of controls is to isolate the effect of a particular 'treatment' (i.e., stressor). This means that controls must be as similar to the stressed sites as possible, except for the stressor of interest (i.e., nutrients). If stressed sites have a suite of other stressors present, then so should the control sites. Other sampling designs can also be considered if they can reduce uncertainty and account for variability and covariates better than BACI designs.

ACCOUNTING FOR ECOSYSTEM PROCESSES

Ideally, site selection should also consider abiotic and biotic mediation of energy sources and flow. Accordingly, Dodds (2007) discussed the importance of trophic state, which are primarily heterotrophic or autotrophic in influencing nutrient utilization and therefore different thresholds of impacts. Even within such zones, site-specific conditions greatly influence stream metabolism and algal growth (Dodds 2007, Hill et al. 2009), and other characteristics. For example, flood frequency also influences periphyton growth and accumulation and all else being equal, eutrophication effects will likely be stronger under stable flow regimes (i.e., intermediate disturbance hypothesis) (Biggs, 2000).

MODIFYING AND EXPANDING ECOLOGICAL RESPONSES

UDWQ has identified several structural and ecological responses that can potentially be used to quantify structural and functional ecological responses. These responses have been evaluated on a regional basis to develop response thresholds. Particularly if used in combination, these indicators will improve the accuracy of nutrient-related assessments and provide information that can inform subsequent site-specific assessments. However, in some cases site-specific study designs should consider augmenting these indicators with ecological responses that can potentially improve insight into site-specific management objectives.

While the need to evaluate the site-specific importance of responses is true for all indicators, this is particularly true for biological data. There is a wealth of available high quality algal and macroinvertebrate data already collected by DWQ. This includes the extensive database at the USU WCMAFE Bug Lab (www.cnr.useu.edu) 'bug lab'. Most of these data can be used in the development and refinement of site-specific nutrient criteria metrics. Relatively few taxa will occur at site-specific sites compared to the region wide taxa pool. It is well known that even species within the same genus can respond to environmental stressors in markedly different ways (e.g., Macan 1963, Downes 2010), therefore, generalized taxonomy is not very useful and is a poor guide to ecology. A thorough knowledge of the life histories and ecologies of these few taxa, particularly indicator taxa is necessary.

Many metrics that are already in use in assessments can be modified and along with newly suggested metrics, can be combined for site-specific assessments. Many more potentially useful metrics should be examined and considered and are discussed in more detail in Appendix B. These metrics can include metrics based on species traits (e.g., functional feeding groups, life histories, voltinism, habitat associations, etc. (see Statzner and Beche 2010 and Vieira et al. 2006)) and bioassessment programs are beginning to reconsider and integrate how environmental stressors can affect these species traits. Examples of potential metrics include:

- Algal taxa richness
- Macroinvertebrate taxa richness
- Algal and macroinvertebrate assemblage structure
- Total algal (periphyton) biomass
- Total macroinvertebrate biomass (abundance corrected)
- Individual indicator algal biomass
- Ex. *Cladophora* sp. and *Didymosphenia geminata*
- Individual indicator macroinvertebrate biomass
    - e.g., scraper taxa or individual snail taxa
    - ex. *Potamopyrgus antipodarum* and *Corbicula* sp.
- Functional feeding group ratios (both taxa and biomass based)
- Secondary production of key indicator species of different functional feeding groups

Changes in biomass and production are much better indicators of site-specific responses to nutrients than abundance measures, which is discussed in more detail in Appendix B.

Macroinvertebrate Collections

Site-specific investigations should consider whether regional biological collection methods should be modified to better account for important site-specific characteristics. It has long been acknowledged that (e.g., Barbour et al. 1999) different floral and faunal assemblages respond over different spatial-temporal scales, and therefore respond to different scales of covariates. However, the framework of bioassessments has not yet fully utilized this realization. This is important because we know that there are large amounts of variation in the structure of benthic communities that could never be fully accounted for by regional relationships. On a site-specific basis there are several smaller-scale variables can cause dramatic shifts in community composition within a 1-m span of riffle. Although, it is often assumed that large composite samples account for this kind of variation through "homogenizing the assemblage", this is not always the case (Marshall 2008, q.v., Appendix B) and site-specific study designs should carefully consider the pros and cons of composite samples.

The patch dynamics of species within a single riffle are affected by many variables such as the accumulations of fine and coarse detritus, the distribution and relative abundance of fine and coarse substrata, algal patch dynamics, and very local dynamics of flow (e.g., Minshall 1984, Barton and Smith 1984, Newbury 1984, Hansen et al. 1991, Hart and Finelli 1999, Lancaster and Downes 2010). The relative importance of these sources of variation, among others, will depend on site-specific circumstances and the specific ecological response. The important point with respect to study designs is that local scale habitat characteristics should all be considered when establishing site-specific collection methods for any of the ecological responses. For instance, near substrata water velocity is known to have pervasive effects on benthic community structure (e.g., Hansen et al. 1991, Hart and Finelli 1999, Lancaster and Downes 2010), which could be accounted for with a flow-stratified sampling regimen to prevent statistical confounding of the variables VELOCITY from treatment effects (i.e., SITE; q.v., see Appendix Bfor details).

Ultimately, accounting for local-scale covariables along with species traits and other response variables (i.e., stream metabolism, water chemistry and habitat assessments) should make the invertebrate assessment results more congruent with other lines of evidence. Reducing the uncertainty of responses associated with covariables, should make all aspects of the assessment more informative by strengthening our understanding of the causal connections between nutrients or important covariables and these responses.

We further expand on how improved sampling design, field sampling and laboratory analysis can be accomplished for the response variables focusing on macroinvertebrate assemblages in Appendix B. These strategies can also be modified for diatoms/algae or other response biota.

Temporal Variation

Spatial variation is often combined or confounded with temporal variation at several time scales. These are discussed in greater detail in Appendix B, but there are several ways in which temporal variation should be accounted for (or controlled) in site assessments. Climatic variation is represented by changes among years in terms of temperature, rainfall, and other atmospheric characters (cloud cover, atmospheric pressure, storm frequency, etc.). It is best dealt with through sampling over a timeframe of several years. A BACI or a BACIPS design (e.g., Osenberg et al. 1994) is useful because the differential response of the study site and the internal reference ("Control") can be used to estimate the degree of impact in all but the most extreme situations.

Seasonal variation can also complicate the interpretation of site assessments. For short-lived organisms (e.g., algae, bacteria, some invertebrates), frequent sampling is required to attain a reasonable estimate of the median annual condition. Furthermore, frequent sampling helps to reduce the risk of uncertainty by preventing ephemeral shifts in production or taxonomic composition from being mistaken for chronic or acute problems. For longer lived organisms (e.g., macroinvertebrates or fish), the effects of seasonal variation are observed through the life cycles and life histories of animals; through emergence, or migrations. Isolating the effects of nutrients is difficult when a group of species are present one year and then absent the next because of temperature-sensitive life-histories (e.g., Sweeney 1984). The most common ways of reducing seasonal variation with longer-lived organisms is to sample the same time period(s) each year. For macroinvertebrates, it is best to avoid time periods when many species are emerging from the system (spring). Thus, winter months would be ideal—other than the interference of snow and ice. Although some insect species emerge in autumn, most of these species have multivoltine, asynchronous larval development and will not affect the richness or diversity as severely as the large emergences of univoltine insects occurring shortly before and shortly after spring runoff. These too can be accounted for through adequate sampling design (Osenberg et al. 1994).

Patch dynamics (Townsend 1989 and Winemiller et al. 2010) are not only an important source of spatial variation in site assessments; they are also a source of temporal variation (seasonal and annual). For example, the structure and function of macroinvertebrate assemblages are intimately tied to flow (e.g., Hart and Finelli 1999), but the ideal refugia for a species may change over time. This can occur because of ontogenic shifts in the speices' life history or simply because of natural physical changes in streams (e.g., spring discharge vs. late summer discharge). Thus, sampling at the exact same geographical location from one sampling period to the next may introduce unwanted variability because the habitat has functionally changed since the previous sampling period. The flow-stratified-systematic-sampling regimen described in Appendix B and alluded to elsewhere in this chapter (i.e., analysis of covariance, Milliken and Johnson 2002) circumvents this source of variation. Models based on regional data do not control for these types of temporal variation, leaving managers to assume all variance is due to stochastic processes or ecological impairments.

Application to Site-Specific Investigations

While broadly applicable, the considerations discussed in this chapter are most relevant to Utah's nutrient reduction strategy as the emphasis shift from regional stressor-response relationships to site-specific applications. With respect to nutrients, there are several management objectives that these investigations might address.

In most cases, the ultimate goal may be derivation of site-specific numeric nutrient criteria. However, other sources of uncertainty may need to be resolved first. For instance, it may be necessary to evaluate regional responses and associated thresholds to ensure that the water quality goals are appropriate for the site of interest. Without appropriate goals, it is impossible to derive criteria that are protective of their associated uses. In other circumstances, the relative role of nutrients and other human-caused may be of greater interest because it may be more effective to restore ecological responses if restoration plans consider the breadth of stressors that are causing ecological problems. In practice, these objectives, among others are not mutually exclusive. Several may be important and the relative emphasis that is placed on each will be situation dependent. This section reviews the concepts previously discussed in the chapter, with an emphasis on how these concepts may apply to site-specific investigations.

SELECTING APPROPRIATE RESPONSES

Response metrics that are most appropriate for quantifying the chemical, physical and biological integrity of the site need to be selected before detailed study designs can be established. Candidate indicators may include those evaluated in this report or others as yet to be determined. For instance, the characterization of sources and sinks of organic matter may be more important at sites with low levels of DO than sites with other nutrient-related water quality problems. Similarly, at sites where daily fluctuations of DO are low it may be difficult or even impossible to obtain measures of whole stream metabolism, so alternative approaches may be more appropriate. Similar considerations apply to measures of important covariates. For instance, channel shading may also be of greater importance for GPP or algae indicators, whereas temperature and substrate size may be more important considerations for macroinvertebrate metrics.

Response metrics may also be modified to better define management objectives. For example circumstances may arise where ongoing investigations reveal that ecological response thresholds are not attainable due to irreversible alterations in hydrology or habitat. In these circumstances, site-specific investigations may focus on identifying indicators that represent the best attainable condition for the stream. Appropriate indicators may be determined based on the ecological outcomes that would be expected if remediation efforts were implemented. These interim indicators can then be reevaluated and if necessary changed through adaptive monitoring and assessment.

SELECTING APPROPRIATE STUDY SITES

In circumstances where reference conditions are unattainable, control sites should be selected that are representative of the best attainable conditions. In some cases, the best site may be a specific segment of the same stream that is in appreciably better condition as long as the segments to be compared are of similar habitat along the river continuum. Wherever possible, several comparable reference or control sites should be selected to estimate the influence of remaining sources of variability.

DEVELOPMENT OF DETAILED SAMPLE ANALYSIS PLANS

Once indicators and sites have been selected a detailed Sample and Analysis Plan (SAP) can be created that will best reduce uncertainty. The SAP will define data quality objectives that take into consideration the many sources of variability discussed in this chapter. Specific collection methods that best account for important sources of variation will need to be described. Where possible, the analytical methods that will be used to derive goals should also be identified *a priori* because this will help determine, the scope, sample frequency and collection efforts that may be required. The SAP should also identify a process that allows the SAP to evolve. Finally, the SAP should specifically identify the roles and responsibilities for all parties involved in the investigation, including PIs, laboratories and field personnel.

Chapter 11

USING QUAL2K MODELING TO SUPPORT NUTRIENT CRITERIA DEVELOPMENT AND WASTELOAD ANALYSES IN UTAH

B.T. Neilson[1], A.J. Hobson[1], N. vonStackelberg[2], M. Shupryt[2], J. Ostermiller[2]

1. Civil and Environmental Engineering
Utah Water Research Laboratory
Utah State University

2. Utah Department of Environmental Quality
Division of Water Quality

NOTE TO TECHNICAL TEAM: WE'LL BE REVISING THIS REPORT TO MORE DIRECTLY RELATE TO NEXT SITE-SPECIFIC STANDARDS. FEEL FREE TO READ IF IT IS OF INTEREST, BUT WE'RE NOT ASKING FOR A FORMAL REVIEW. ADDITIONAL INFORMATION CAN BE FOUND IN APPENDIX C.

A cooperative data collection and modeling effort between the Utah Department of Environmental Quality (Utah DEQ), Division of Water Quality and Utah State University (USU) began in 2010. The primary objectives of this study were to 1) design a data collection approach appropriate to support the population and calibration of QUAL2Kw models for use in a variety of applications; and 2) develop a methodology for populating and calibrating QUAL2Kw given these data. The intended use of the resulting models was to assist in developing numeric nutrient criteria for the state of Utah and provide a starting point for the development of new waste load allocations (WLAs) for 9 water reclamation facilities (WRFs). The objectives were completed by assisting DEQ in collecting the appropriate data in the reaches below WRFs around the state and using these data to populate and calibrate QUAL2Kw models for each of these study sites.

Background

Over the last few years the state of Utah has been working towards understanding the implications of instituting numeric nutrient criteria. To do this they have initiated a publicly owned treatment works (POTW) nutrient removal cost study [*UDEQ*, 2009], an economic evaluation study [*UDEQ*, 2011], and a nutrient criteria ecological study [*UDEQ*, 2010]. The Nutrient Removal Cost Impact Study, completed in 2009, evaluated the economic impacts of potential new nutrient removal requirements for Utah's POTWs. The study estimated economic, financial, and environmental impacts interrelated with a range of potential nutrient discharge standards for every discharging mechanical POTW in the State and one lagoon system [*UDEQ*, 2009]. The economic evaluation study, which is still in progress, is intended to quantify the economic benefits and costs of implementing nutrient criteria for surface waters in Utah [*UDEQ*, 2011].

When investigating nutrient criteria based on the ecological implications, EPA recommends three types of scientifically defensible empirical approaches for establishing numeric thresholds intended to limit nitrogen/phosphorus pollution: reference condition approaches, mechanistic modeling, and stressor-response analysis [*US EPA*, 2010]. The DEQ is currently investigating all three recommended approaches to establish numeric nutrient criteria. In order to complete two of the recommended approaches the DEQ, along with USU, are investigating the ecological impacts of nutrients on Utah waterbodies using both a stressor response approach combined with the predictive capabilities of the mechanistic modeling approach. To do this a data collection strategy was developed to meet the needs of both recommended approaches. By combining the results of the economic study with the predictive capabilities of the modeling efforts and ecological response information, the proposed instream nutrient criteria can be linked to the expected economic costs of the treatment upgrades as well as forecasting the potential impact of nutrient loading on the ecological health of the downstream waterbodies.

This report covers the general approaches taken in the mechanistic modeling portion of the Nutrient Criteria study for data collection, model population, and calibration/validation. However, it is important to note that the data collection approaches and models are intended to have multiple applications and therefore, have been made very generic in order to support: 1) development of statewide numeric nutrient criteria; 2) development of site-specific criteria for rivers and streams where the statewide nutrient criteria do not appear valid; 3) wasteload analyses to determine water quality based effluent limits (WQBEL), and 4) determination of TMDL endpoints. Details regarding the associated ecological measures and reference condition information can be found at http://www.nutrients.utah.gov/nutrient/index.htm.

QUAL2Kw Model

Utah DEQ uses low flow conditions to determine WQBELs for point sources under the Utah Pollution Discharge Elimination System (UPDES) program (*UDEQ*, 2012) due in part to these corresponding with limiting conditions. This led to selecting a model that would be appropriate for these conditions. QUAL2K [*Chapra et al.*, 2004] is a USEPA approved model that has been commonly used in WLAs and total maximum daily loads (TMDLs) (e.g., [*Bischoff et al.*, 2010; *Kardouni and*

Cristea, 2006]), and even development of nutrient criteria [*Flynn and Suplee*, 2011]. This quasi-dynamic, one dimensional instream water quality model includes the dominant processes of concern within Utah waters, predicts the required water quality variables, and is feasible to populate and calibrate given the limited data available in most waterbodies of the state. In order to understand the associated daily minimum and maximum instream concentrations, the model provides a 24 hour diel response in water quality given an appropriate or representative 24 hour weather pattern. QUAL2Kw [*Pelletier and Chapra*, 2008], a sister model to QUAL2K developed within the state of Washington, built in additional functionality (e.g., automatic calibration algorithms) into QUAL2K based on their identified needs. With the anticipation of having some similar needs as Washington and the possibility of identifying additional needs, Utah DEQ elected to use QUAL2Kw in their instream modeling applications.

Details regarding the version of QUAL2Kw used in this application (version 5.1) are provided within the user's manual [*Pelletier and Chapra*, 2008] and a number of publications [*Cho and Ha*, 2010; *Kannel et al.*, 2007; *Pelletier et al.*, 2006]. In short, the state variables (Table 1) include the macro nutrients (C, N, and P) of interest and the critical nutrient species (e.g., inorganic P, nitrate, and ammonia) in surface waters.

Using the same notation as that of Table 1, the QUAL2Kw composite or calculated variables are (Pelletier and Chapra, 2008):

Total Organic Carbon (mgC/L):

$$TOC = \frac{c_s + c_f}{r_{oc}} + r_{ca} a_p + r_{cd} m_o \quad \textbf{(1)}$$

Table 11.1. QUAL2Kw State Variables (taken directly from [*Pelletier and Chapra*, 2008]).

Variable	**Symbol**	**Units***
Conductivity	s_1, s_2	µmhos
Inorganic suspended solids	$m_{i,1}$, $m_{i,2}$	mgD/L
Dissolved oxygen	o_1, o_2	mgO_2/L
Slow-reacting CBOD	$c_{s,1}$, $c_{s,2}$	mg O_2/L
Fast-reacting CBOD	$c_{f,1}$, $c_{f,2}$	mg O_2/L
Organic nitrogen	$n_{o,1}$, $n_{o,2}$	µgN/L
Ammonia nitrogen	$n_{a,1}$, $n_{a,2}$	µgN/L
Nitrate nitrogen	$n_{n,1}$, $n_{n,2}$	µgN/L
Organic phosphorus	$p_{o,1}$, $p_{o,2}$	µgP/L

Variable	Symbol	Units*
Inorganic phosphorus	$p_{i,1}$, $p_{i,2}$	µgP/L
Phytoplankton	$a_{p,1}$, $a_{p,2}$	µgA/L
Detritus	$m_{o,1}$, $m_{o,2}$	mgD/L
Pathogen	x_1, x_2	cfu/100 mL
Generic constituent	gen_1, gen_2	user defined
Alkalinity	Alk_1, Alk_2	$mgCaCO_3/L$
Total inorganic carbon	$c_{T,1}$, $c_{T,2}$	mole/L
Bottom algae (a_b in the surface water layer), biofilm of attached heterotrophic bacteria (a_h in the hyporheic sediment zone for the Level 2 option)	a_b,a_h	gD/m^2
Bottom algae nitrogen	IN_b	mgN/m^2
Bottom algae phosphorus	IP_b	mgP/m^2

* mg/L ≡ g/m^3

Total Nitrogen (µgN/L):

$$TN = n_o + n_a + n_n + r_{na} a_p \tag{2}$$

Total Phosphorus (µgP/L):

$$TP = p_o + p_i + r_{pa} a_p \tag{3}$$

Total Kjeldahl Nitrogen (µgN/L):

$$TKN = n_o + n_a + r_{na} a_p \tag{4}$$

Total Suspended Solids (mgD/L):

$$TSS = r_{da} a_p + m_o + m_i \tag{5}$$

Ultimate Carbonaceous BOD (mgO_2/L):

$$CBOD_u = c_s + c_f + r_{oc} r_{ca} a_p + r_{oc} r_{cd} m_o \tag{6}$$

Additionally, the model provides the ability to predict the associated biological effects of various nutrient concentrations since photosynthesis, respiration, and death of phytoplankton and bottom algae are included within the model. As the version of QUAL2Kw applied within this study is quasi-dynamic, it provides the ability to deal with steady flow, but does allow for non-uniform flow. This means that while the flow conditions cannot change over time, they can vary longitudinally downstream due to point or distributed inflows or abstractions.

Given the capabilities of this version of QUAL2Kw, there are environmental conditions that are suited for this type of modeling approach. The time period over which this model should be applied require that 1) stream conditions are completely mixed since the model assumes all model elements are completely mixed, 2) boundary condition concentrations can be approximated by consistent 24 hourly values; 3) distributed flows are constant, 4) point inflows follow a consistent diel pattern or are constant, and 5) weather conditions over the simulation period have a consistent diel pattern.

Study Site Locations

Nine sites were selected for the nutrient criteria ecological study and represent the different types of receiving waterbodies around the state of Utah (Figure 11.1). Using these sites as a representative sample of the state's waterbodies, the QUAL2Kw, ecological stressor-response, and reference condition findings will be used to extrapolate information regarding possible ranges of nutrient criteria for the remaining state waters [*UDEQ*, 2010]. The selected sites (Table 11.2) are located within different order streams with varied background water quality, surrounding land uses, and amounts of wastewater effluent that have been treated to different levels. The sections studied were those influenced by WRF effluents since these areas generally have enhanced nutrient loads. More detail regarding each site (e.g., location, study reach length, etc.) are provided in a separate report in preparation by the Division of Water Quality that evaluates structural and functional responses to nutrients. Detailed information about unique sampling requirements associated with each site and the specific information regarding model population, calibration, and validation are provided within the QUAL2K modeling files and site specific model documentation provided to Utah DEQ as project deliverables.

Figure 11.1. Study site locations within the state of Utah.

Table 11.2. Study site locations, water reclamation facilities, and dates sampled within the state of Utah.

Waterbody	Facility	Dates Sampled
Box Elder Creek	Brigham City WRF	Aug. 9 - 11, 2010
San Pitch River	Fairview City WRF	Aug. 2 - 5, 2010 Oct. 11 - 13, 2010
San Pitch River	Moroni City WRF	July 28-30, 2010
Weber River	Oakley City WRF	Aug 23-26, 2010
Price River	Price River Water Improvement District	Aug. 30 - Sep 1, 2010
Dry Creek	Spanish Fork City WRF	July 23 - 26, 2010
Silver Creek	Snyderville Basin-Silver Creek Water Reclamation Facility	July 20 - 22, 2010 Sep. 30 - Oct 4, 2010 Aug 22-30, 2011
Malad River	Tremonton City WRF	Aug. 13 - 16, 2010
Little Bear River	Wellsville Lagoons	Sept 10-13, 2010

Project Results

Since the key objectives in this project were to develop the appropriate data collection methodologies to support QUAL2Kw model population and calibration, this report provides general information regarding the field data requirements, approaches to model population using these data, strategies used in model calibration, and the steps required for model validation (if these data sets exist).

Supporting field data

Data must be collected at 3 general locations for instream modeling. The beginning of the study reach (also called the headwater or upstream boundary condition), inflows/outflows (point sources or tributary inflows and diversions/abstractions), and at least one location downstream for model calibration. The data types required at these locations will vary and are discussed below. For the 2010 data collection efforts, data collection at each location spanned a 2 day period during low flow conditions.

Figure 11.2 shows a generalized schematic used within the 2010 data collection efforts. Data for the modeling efforts were gathered at Station B (headwater/upstream boundary condition), Station C (wastewater treatment plant effluent before it enters the stream), Station D (at a location where the stream and point source effluent was completely mixed), and Station E (the calibration location downstream at the end of the study reach). Information gathered at Station A was only used in the ecological portion of the study. If a tributary entered the modeling reach, data were also collected at T1. Similarly, if a diversion was present, the quantity of water leaving the system was determined.

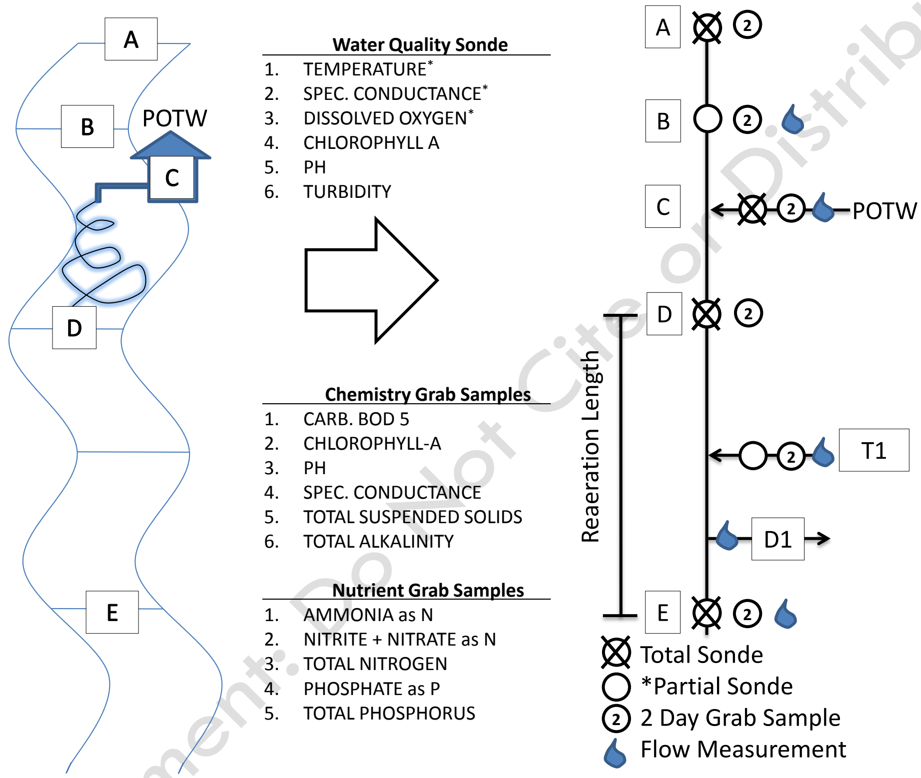


Figure 11.2. Generalized data collection locations within the 2010 sampling efforts. Required locations of flow measurements and the multi-parameter water quality sondes are also shown. Since 2010 modeling efforts used Station B as the headwater location, the information for chlorophyll-a and pH is taken from Station A.

For the 2010 data collection, the location of the completely mixed conditions downstream of the WRF was determined by measuring specific conductance or temperature across the channel to determine where uniform conditions existed. In some cases where the differences in temperature and/or specific conductance were too small, rhodamine WT was used as a visual indicator. To support and integrate these efforts with the ecological study needs the distance between Station D and E was estimated using methods described in Grace and Imberger [2006] which designate the optimum distance between stations for calculating open water metabolism using the single station method (Eqn. 7).

$$X = 0.693 \cdot \frac{v}{k_a} = \frac{v^{0.33}}{0.0137 \cdot \left(D^{-0.85}\right)} \qquad \textbf{(7)}$$

Where, X = optimum station distance (km), v = velocity (cm s^{-1}), D = depth (cm), and k_a = reaeration coefficient for oxygen (d^{-1}).

As discussed later, we found this method to result in distances that were in general too short to meet the diverse needs of this study and often times did not include the compliance point for WLAs.

The information necessary at each of these stations is dependent on whether it is the headwater location, a load, or a diversion. Water quality models require an understanding of both the water balance and the mass balances for each constituent modeled. Flow measurements may be required at all stations in order to establish a water balance. Water quality information is not, however, required for diversions or abstractions since the mass loss will be a function of the instream concentrations predicted by the model and the volume of water taken out that is specified by the user.

The specific water quality constituents measured at each station and the frequency they were collected are detailed within Table 3. There are a number of constituents that were measured using multi-parameter sondes (e.g., temperature, dissolved oxygen, conductivity) at five minute increments over each of the two day sampling period. Grab samples of most constituents requiring laboratory analyses were gathered each day and usually in the mid-morning. Benthic algae sampling was only conducted once at some point in time close to the study periods. A number of constituents within this list, indicated by a * in the table, were not sampled directly and had to be estimated. The appropriate values for modeling were estimated using the relationships between measured constituents and model variables as described below. Additional data types that could be collected that would be useful in the modeling include a measure of sediment oxygen demand, total organic carbon, and volatile suspended solids. None of these measures were completed in the 2010 data sets.

Data to characterize each site is additionally necessary to support model population or calibration. Table 4 provides a list of the data types requiring collection, some procedural information, locations where these data are required within or near the site, and the utility of the data in the context of the modeling effort. A number of these data types are collected within routine Utah's Comprehensive Assessment of Stream Ecosystems (UCASE) surveys based on protocols adapted from the USEPA [2007]. It is important to note that all locations where data are collected must have GPS coordinates established for documentation purposes.

Model Population

Once the data have been collected, they must be translated from observations to model inputs. The model state variables (Table 1) can be related to measurements as follows [taken directly from Pelletier and Chapra, 2008]:

Conductivity = s = COND **(8)**

ISS = m_i = TSS – VSS or TSS – r_{dc} (TOC – DOC) **(9)**

Dissolved Oxygen = o = DO **(10)**

Organic Nitrogen = n_o = TKN – NH4 – r_{na} CHLA or **(11)**

n_o = TN – NO2 – NO3 – NH4 – r_{na} CHLA

Ammonia Nitrogen = n_a = NH4 **(12)**

Nitrate Nitrogen = n_n = NO2 + NO3 **(13)**

Organic Phosphorus = p_o = TP – SRP – r_{pa} CHLA **(14)**

Inorganic Phosphorus = p_i = SRP **(15)**

Phytoplankton = a_p = CHLA **(16)**

Table 11.3. Water quality constituents sampled and the frequency of sampling for QUAL2Kw modeling.

Multi-Parameter Sonde Data	**Abbreviation/QUAL2Kw Units**	**Frequency**
Water Temperature	Temp (C)	5 min samples
Specific Conductance	COND (mhos)	5 min samples
Dissolved Oxygen	DO (mgO_2/L)	5 min samples
pH	pH	5 min samples
Chlorophyll a	CHLA (gA/L)	5 min samples
Turbidity		5 min samples
Laboratory Analysis		
5-Day Soluble Carbonaceous BOD, sCBOD5		1 each day
Total Nitrogen	TN (gN/L)	1 each day
Ammonia Nitrogen	NH_4 (gN/L)	1 each day
Nitrate+Nitrite Nitrogen	NO_3 (gN/L)	1 each day
Total Phosphorus	TP (gP/L)	1 each day
Soluble Reactive Phosphorus	SRP (gP/L)	1 each day
Volatile Suspended Solids*	VSS (mgD/L)	1 each day
Total Suspended Solids	TSS (mgD/L)	1 each day
Alkalinity	ALK ($mgCaCO_3$/L)	1 each day
Chlorophyll a	CHLA (gA/L)	1 each day
Dissolved Organic Carbon, DOC	DOC (mgC/L)	1 each day
Dissolved Organic Phosphorus, DOP*		1 each day
Dissolved Organic Nitrogen, DON*		1 each day
Benthic Chl-a		1 per sampling time period
Benthic AFDM		1 per sampling time period
Benthic TP		1 per sampling time period
Benthic TN		1 per sampling time period
Benthic TOC		1 per sampling time period
SOD#		1 per sampling time period
TOC#	TOC (mgC/L)	1 each day

* = not gathered or required estimation for QUAL2Kw
# = data that would be useful in model population/calibration but were not directly measured in these efforts

Table 11.4. Site characterization data types.

Data Type	Procedure	Locations	Reasoning
Average Cross Sectional Velocity*	See methods provided within Data Collection and/or UCASE SOP. Information from HEC-RAS modeling applications can also be extracted to supplement data collected.	Station D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of velocity in different reaches to compare with the predicted velocities. This can be used with the depth and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Cross Sectional Depth*	See methods provided within Data Collection and/or UCASE SOP.	Station D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of depths in different reaches to compare with the predicted depths. This can be used with the velocity and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Channel Bottom Width	Bottom width estimates were calculated using side slope, average depth, and top width values in the formula: Top Width – Depth x 1/tan(radians(°SSLEW)) – Depth x 1/tan(radians (°SSREW)), where width and depth are in meters and side slope is in radians in the form of Run/Rise.	Station D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input.
Channel Bottom Slope	See methods provided within UCASE SOP.	Should estimate bottom slope from beginning to end of study reach at 10% increments of total reach length and/or when changes in bottom slope are observed.	Model input.
Channel Side Slope	See methods provided within UCASE SOP.	Station D, E, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input and can be used to calculate bottom width from measured top widths.
Weather data	Onsite weather station or nearest Mesowest Station.	Near study site would be most appropriate and 15-30 minute data are preferred.	Wind speed, air temperature, shortwave solar radiation, humidity/dewpoint temperature are all used within the model as forcing information. Precipitation data shows whether there was significant rainfall in the area that would influence instream flows.
Tracer Study	Inject tracer at Station B or C and measure response at Station E. Can also use HEC-RAS model if available.	Measure tracer response at Station E, but additional locations along the study reach would be beneficial to capture heterogeneity.	Provides information regarding average travel time through system and can be used in calibration of hydraulic parameters (e.g., Manning's roughness coefficient).
Substrate type*	See methods provided within Data Collection and/or UCASE SOP.	Information should be gathered at cross sections in subreaches that represent the variability in substrate types.	Provides a method to approximate the Mannings roughness coefficient and determine fraction of bottom substrate appropriate for bottom algae.
Shading*	See methods provided within Data Collection and/or UCASE SOP.	Information should be gathered at locations that represent the variability in shading.	Model input. If riparian or topographic shading drastically influences instream temperatures, estimates of the shading % for each hour of a day will be necessary to scale the incoming shortwave solar radiation.

$$\text{Detritus} = m_o = \text{VSS} - r_{da}\,\text{CHLA or } r_{dc}\,(\text{TOC} - \text{DOC}) - r_{da}\,\text{CHLA} \quad \textbf{(17)}$$

$$pH = \text{PH} \quad \textbf{(18)}$$

$$\text{Alkalinity} = Alk = \text{ALK} \quad \textbf{(19)}$$

While a number of these relationships are straightforward, it is important to realize that the typical Organic N and Organic P measurement cannot be directly compared to the Organic N and Organic P QUAL2Kw predictions. As shown in equations 11 and 14 above, the QUAL2Kw versions of organic N and P only represent the dissolved and detritus portion of each organic nutrient pool since the portion associated with the live algae are subtracted out. It is also important to note that detritus (Eqn 17) only contributes to the carbon budget and does not influence other nutrient pools.

Given the data available from the 2010 sampling, we needed additional methods based on some assumptions or established equations to calculate the variables necessary for model population and calibration. These included the need to convert $sCBOD_5$ measurements to the sCBOD ultimate values required within QUAL2Kw (Eqn 20).

$$\text{sCBOD ultimate} = c_f \text{ or } c_s = \text{sCBOD}_5/(1\text{-exp}(\text{-k}_d\ (5\text{days})) \quad \textbf{(20)}$$

Further, since we did not have direct measures of VSS or ISS in the 2010 data sets, we had to come up with methods and a logic tree to estimate these values for model population (Figure 3).

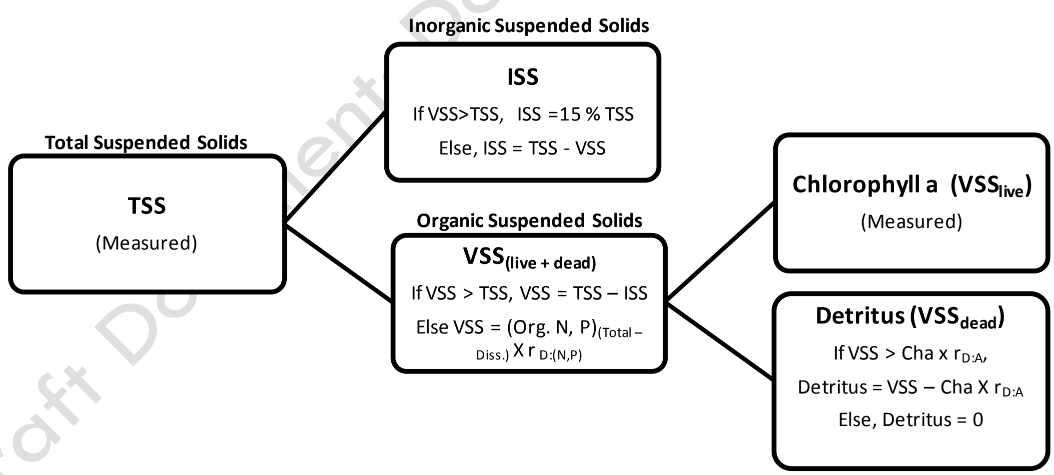


Figure 11.3. Logic used in estimating VSS and ISS from TSS, followed by logic for estimating detritus from VSS.

To populate the model, information regarding the reach, initial conditions, headwater conditions, weather data, point sources and distributed sources must be provided (Table 5). More specifically, observations from the headwater location and any point flow (inflow or abstraction) or distributed flows must be entered into the model framework. Any flow information provided for these locations must be a representative value for the entire modeling period. The necessary sampling frequency of specific water quality data is dependent upon whether it is a point source or headwater (Table 6). The other forcing data

Table 11.5. General information required for QUAL2Kw model population.

QUAL2Kw Sheet	**Information Required**
Reach	
	Reach segmentation
	Hydraulic characteristics
	% suitable substrate
	Bottom algae % cover
	SOD
	Thermal properties
Initial Conditions	
	Constituent concentrations (See Table 6)
Headwater Data	
	Average flow
	Constituent concentrations (See Table 6)
Weather Data (hourly average values)	
	Air temperature
	Dewpoint temperature
	Solar radiation
	Shading
	Cloud cover
	Wind speed
Point Sources	
	Average flow
	Constituent concentrations (See Table 6)
Distributed Sources	
	Average flow
	Constituent concentrations (See Table 6)
Rates	
	Primarily set in calibration. See Model Calibration section below.

TABLE 11.6. MODEL INPUT CONSTITUENT CONCENTRATIONS REQUIREMENTS AND THE ASSOCIATED OBSERVED DATA USED IN POPULATION OF QUAL2KW.

		Point Source	**Headwater**	**Distributed Inflow**
Model Parameter	**Data Collected**	Mean + Range/2 or 2 Day Mean	Hourly Average or 2 Day Mean	Average
Alkalinity	Total Alkalinity	X	X	X
$sCBOD_{ultimate}$	$sCBOD_5$	X	X	X
Specific Conductivity	Specific Conductivity	X	X	X
Detritus (POM)	(Org - Diss. N, P) X r(POM/N, P)	X	X	X
Dissolved Oxygen	Dissolved Oxygen	X	X	X
Inorganic Phosphorus (SRP)	Inorganic Phosphorus (SRP)	X	X	X
Inorganic Solids	TSS - VSS	X	X	X
NH_4-Nitrogen	NH_4-Nitrogen	X	X	X
NO_3-Nitrogen	NO_3-Nitrogen	X	X	X
Organic Nitrogen	TN - (NH_4) - (NO_3 + NO_2)	X	X	X
Organic Phosphorus	TP - Inorg P	X	X	X
pH	pH	X	X	X
Phytoplankton	Chlorophyll a	X	X	X
Water Temperature	Water Temperature	X	X	X

required by the model is meteorological information which includes hourly average air temperatures, wind speeds, and dewpoint temperatures from a nearby, representative weather station. Shortwave solar radiation can be estimated automatically within the modeling framework, however, if using these estimates, hourly cloud cover values would be required. In the 2010 modeling efforts, we instead used actual shortwave radiation observations from a local source.

When populating the models, censored data, or concentrations that are below the analytical detection limits (i.e., non-detects) commonly occur. Within the 2010 modeling, non-detects were assigned a concentration of half of the detection limit. More accurate statistical analysis of limited amounts of censored data should be investigated. The detection limits associated with key parameters are detailed within Table 11.7.

Table 11.7. Detection limits for constituents based on the procedures applied within specific laboratories.

Constituent	Laboratory	Analytical Detection Limit (mg/L)
TN, TP	Baker Lab – USU	0.0057
TDN, TDP	Baker Lab – USU	0.0025
NO_3+NO_2 -N	Baker Lab – USU	0.0006
NH_4 -N	Baker Lab – USU	0.00395
PO_4 -P	Baker Lab – USU	0.0008
$sCBOD_5$	Utah DEQ Laboratory/AWAL	3/5
Chlorophyll a	Utah DEQ Laboratory	0.0007
Specific Conductance	Utah DEQ Laboratory	2 (uS/cm)
Total Suspended Solids	Utah DEQ Laboratory	4
Total Dissolved Solids (180 °C)	Utah DEQ Laboratory	10
Turbidity	Utah DEQ Laboratory	0.1 (NTU)

To assist in ensuring model population consistency given the relatively consistent data collection strategies implemented in 2010, we developed two supporting sheets within the QUAL2Kw files delivered to DEQ. A "Data Input" and "Addt Info" sheet provides a number of tables that can be populated with observations, and this information automatically populates the QUAL2Kw sheets. Further, these sheets facilitate some of the additional calculations that were completed and suggested in future applications (described further below).

Most information within the Rates Sheet was not changed at all or was adjusted in model calibration (described further below). However, specific values of some parameters were established within the 2010 modeling efforts that may be appropriate for other Utah model applications. First, we measured CBOD decomposition rates (k_d) by taking 6 samples from the Silver Creek WRF effluent. These samples were analyzed in triplicate resulting in 18 total measurements of 30 day CBOD using methods detailed in Environmental Protection Division [1989]. The resulting data were analyzed using a Nonlinear Least Squares Method and the Thomas Method (Table 11. 8). Given that Chapra [1997] reports values ranging 0.05-0.1 d^{-1} at 20°C for waste streams treated using activated sludge, we assumed the average value of 0.103 d^{-1} was an appropriate value for all the 2010 study sites and this value was not varied in calibration. Further, this value was used to convert any measured concentrations of $sCBOD_5$ to the sCBOD ultimate values required by the model (Table 11. 6). We do, however, suggest that a Utah specific number be established for the dominant wastewater treatment types of activated sludge and membrane mechanical treatment as well as for lagoon systems.

Table 11.8. CBOD decomposition rate statistics based on samples from Silver Creek WRF effluent.

	NLS Method	Thomas Method
	k_d, 1/d	k_d, 1/d
Min	0.095	0.076
Max	0.125	0.124
Mean	0.103	0.096
StDev	0.011	0.012
95% CI	0.013	0.013

The thermal properties of the Silver Creek substrate were also measured since these dictate the rate of heat exchange between the water column and the sediments. While there are a number of values reported within the QUAL2K and QUAL2Kw manual, it can be important to have site specific thermal properties. The thermal property values based on measurements from Silver Creek with a sandy-gravel substrate were a thermal diffusivity of 0.72 $mm^2\ s^{-1}$ and a thermal conductivity of 2.25 $W\ m^{-1}\ k^{-1}$.

Model Calibration

Calibration within this effort consisted of a number of manual calibration steps followed by autocalibration using the genetic algorithm within QUAL2Kw. The data used in calibration included hydraulics data (longitudinal depths, velocities, and travel time) and water quality data (Table 11. 3) including the mean, minimum, and maximum values at each calibration location (only station E for the 2010 effort). For those data types where only 2 samples were taken, the minimum and maximum values were not always representative of the daily variability and only provided an understanding of the range at these sampling times.

MANUAL CALIBRATION STEPS

A number of manual calibration steps and or checks were identified to ensure that the model was representing the system well based on site specific data. These steps are key since they ensure that the foundational model components (e.g., flow balance, volumes, and instream temperatures) are correct before moving onto the more interconnected mechanisms associated with nutrient cycling.

Flow Balance/Hydraulics - To ensure that the representation of the hydraulics was appropriate, a number of steps were taken and many data types must be considered. Initially, to make sure discharge matched empirical observations, predicted values were compared to measured values. If the values differed, it could have been due to inflows or outflows from unknown sources or from groundwater exchanges. Although the reaches in these studies did not have significantly large differences, in some cases it may be necessary to incorporate a distributed inflow or abstraction to represent groundwater influences.

Specific conductance values were used in a number of different ways. Predicted values were compared to observed specific conductance values including the diel fluctuations. Since specific conductance is a measure of relatively conservative dissolved species, if predictions did not match the observations, this could indicate the presence of unknown inflows and may suggest the need for additional time in the field determining the source of the inflow.

Travel times within the study reach are dependent on having the channel geometry, water depth, and velocities correct. After data collection efforts were completed, we conducted a tracer study using either salt or rhodamine WT to provide data regarding travel times within the study reaches. Because Manning's equation is used to route the water through the study reach, additional information must be provided at a subreach scale about bottom widths, side slope, channel bottom slope, and Manning's roughness coefficient. Top widths, side slopes, and bottom slopes are measured at consistent increments along the channel. From these data, as described in Table 11. 4, bottom width estimates were calculated using side slope, average depth, and top width values. Once the bottom width, side slope, and bottom slope values were entered into the model and the model was run, the predicted top widths were compared to field-derived data. If necessary, the bottom widths or side slopes were adjusted within reason. While good estimates of water depth and velocity were available at a number of discharge measurement locations, these values were not always recorded. After model setup, where available, predicted water depths and velocities were compared to measurements from locations downstream. At the same time, predicted travel times were compared to those estimated from tracer injection responses at various locations downstream. If necessary Manning's n and possibly bottom slope were adjusted to ensure water depths, velocities, and travel time predictions were similar to observations. Once the hydraulic representation was appropriate, it was necessary to determine if the temperature and ISS predictions were acceptable.

Temperature - First, predicted and observed temperatures at different locations downstream were compared. If predictions were inaccurate, shading data and trial and error approaches were used to adjust the hourly percent shading values. Another consideration was the accuracy of the predicted top widths. This was key because the predicted surface heat flux values are dependent upon the surface area of the air water interface. At times, it may be necessary to revisit the top width predictions to ensure the accuracy of temperature predictions. Additionally, if there were inflows, the temperature of these inflows may have required adjustment if they were not measured in the field.

Inorganic Suspended Solids (ISS) - Second, predicted and estimated ISS concentrations were compared at various locations longitudinally. The settling velocity was adjusted to vary the predicted concentrations. Ensuring these values were correct was important for photosynthesis due to its influence on light penetration. However, the ISS observations were calculated and it was unclear if they were accurate.

Reaeration Rates - Finally, to minimize the number of parameters that are varied in the autocalibration we developed an approach to determine the appropriate reaeration formula to apply within the model and a method of approximating SOD using the dissolved oxygen timeseries collected at each site. To determine a representative reaeration formula, whole stream metabolism methods were applied to estimate gross primary production (GPP) and ecosystem respiration (ER) using the concentrations at most stations where DO was measured within the study reach. As part of this, it is necessary to estimate a reaeration rate (k_a). Various "open-water" methods of determining GPP, ER and k_a have been established including the Delta Method [*McBride and Chapra*, 2005], Night Time Regression [Young et al 2004], and Inverse Method [Holtgrieve et al 2010]. It is possible to select the appropriate method for sites based on recommendations of Aristegi et al. [2009] where the Delta method was found to be best in open canopy and clear conditions (using the point method if data are smooth and the centroid method if data are noisy) and the Night Time method was inappropriate in turbulent reaches and where WRF effluent is dominant and there are highly variable flows. For various sites in Utah, the Inverse Method was found to produce the most consistent results based on estimates for many systems and sites across the state.

To support the QUAL2Kw modeling, the Night Time Regression and Inverse Methods were applied to estimate the GPP, ER, and/or k_a at locations along the study reach. Given the variability of predicted reaeration rates from the formulas included within QUAL2Kw and the associated uncertainty, a number of steps were taken to determine the most appropriate formula. First, we would run QUAL2Kw model using each reaeration formula. These predicted reaeration rates in each reach segment were compared to the k_a values estimated from the metabolism methods and an RMSE was calculated. Next, we determined the most appropriate formula based on the lowest RMSE value and set this within the model. If multiple equations were appropriate, we selected one where all assumptions (e.g., depth and velocity ranges) were met.

As described within Appendix A and B, these steps have been automated within the "Data Input" sheet that USU added for the 2010 modeling. It is important to note that it would be possible to set the reaeration rates based on the values obtained from the metabolism measurements directly, however, this would limit the applicability of the model to predict reaeration under other flow conditions (i.e., due to different velocities and depths) and within reach segments where k_a values were not measured.

<u>Sediment Oxygen Demand</u> - Another significant source of uncertainty in QUAL2Kw modeling, particularly in shallow streams, is the amount of SOD present within each system. While QUAL2Kw has the functionality to estimate SOD based on a sediment diagenesis algorithm, there is often more SOD present than is predicted. The need to prescribe SOD has been associated with the deposition of organic matter outside of the time period of the model simulation (i.e. during snowmelt runoff) and the deposition of coarse particulate organic matter (CPOM) that typically is not captured by standard sampling techniques. This extra SOD became an issue within the Jordan River TMDL [*Stantec Consulting*, 2010] and was addressed through direct measurements of SOD to determine reasonable ranges that would be acceptable within QUAL2Kw modeling. In many cases, however, these types of measurements will not be available due to cost and the personnel requirements to collect them and there is significant variability in the results. We know that the change in oxygen over time is a result of oxygen sources (primary production and reaeration) and oxygen sinks (autotrophic and heterotrophic respiration, BOD, and other oxygen consuming reactions within the water column and sediments). However, when using metabolism methods, the equation describing the change in oxygen is reduced to:

$$dO/dt = GPP + reaeration - ER \qquad \textbf{(21)}$$

where, *GPP* = gross primary production and *ER* = ecosystem respiration.

In this context, *ER* is now a net sink term. If we assume autotrophic respiration approximately equals *GPP* (it may need to be some fraction of *GPP [Jones et al., 1997]*), then any extra oxygen consumption is due to heterotrophic respiration and other oxygen consuming reactions within the sediments and water column. If this value is positive (meaning *ER* is higher than primary production), this provides an estimate of a total SOD (heterotrophic respiration + oxygen demanding reactions within the sediments) and some oxygen demanding reactions within the water column (e.g., BOD decomposition and nitrification). Within QUAL2Kw, it can be assumed that this total SOD value would provide a maximum SOD that could be prescribed within the model. In most cases, the maximum SOD should include the prescribed SOD plus the SOD estimated within the sediment diagenesis algorithm within QUAL2Kw (described within [*Pelletier and Chapra*, 2008]). In these efforts, we assumed that the *ER* minus *GPP* approximation for SOD is appropriate since the streams included in this study are relatively shallow and sediment processes will significantly influence the water column DO response. In larger rivers, it is possible that other processes more significantly influence the water column oxygen responses (e.g. chemical reactions within the water column, phytoplankton, etc.) and these approaches may not be applicable or include more error due to the aforementioned assumptions.

Since SOD measurements were not gathered during the 2010 data collection efforts, we used *ER* values minus the *GPP* estimates at Station E (and at times Station D) to determine a reasonable average and range of SOD values for the portion of the study reach below the WRF. Where appropriate, an average value of SOD was established and set within the model before autocalibration.

AUTOCALIBRATION

With a number of parameters set based on the prior manual calibration steps, the remaining parameters that were appropriate to include in model calibration were autocalibrated. The parameters that should be included in calibration as well as the appropriate parameter ranges were set based on recommendations from Dr. Steven Chapra [*Stantec Consulting*, 2010] and from Bowie et al. [1985] (Table 11. 9). Within the autocalibration, a fitness statistic is evaluated for each state variable as the reciprocal of a weighted average of the normalized RMSE and estimated as follows:

$$f(x) = \left[\sum_{i=1}^{q} w_i\right]\left[\sum_{i=1}^{q}\frac{1}{w_i}\left[\frac{\frac{1}{m}\sum_{j=1}^{m} O_{i,j}}{\left[\frac{1}{m}\sum_{j=1}^{m}\left(P_{i,j}-O_{i,j}\right)^2\right]^{1/2}}\right]\right] \quad \textbf{(21)}$$

Where, $O_{i,j}$ = observed value, $P_{i,j}$ = predicted value, m = number of pairs of predicted and observed values, w_i = weighting factor, and q = number of different state variables (e.g., dissolved oxygen, pH) in a bounded n-dimensional space for $x \equiv (x_1, x_2, ..., x_n) \quad x_k \in [0.0, 1.0]$ (Pelletier et al. 2006).

This tool, allows the coefficient of variation of the RMSE (model results versus observed data) between each constituent along with appropriate, individual weighting factors (Table 11. 10), to be summarized in a single value that the genetic algorithm seeks to maximize by adjusting all desired parameters.

The constituents included in the fitness statistic for the 2010 modeling efforts heavily weighted DO average, minimum, and maximum values at Station E as indicted by a weighting factor of 5 and were established via discussions with Greg Pelletier and Nick von Stackelberg. The preliminary calibration parameters for each study site were established by the autocalibration algorithm and are outlined within each model and the associated documentation delivered to UDEQ.

Table 11.9. Appropriate ranges (Min Value and Max Value) of parameters for QUAL2Kw modeling with the "Value" column showing the default value used. The "Auto-Cal" column indicates if a parameter was autocalibrated in the 2010 modeling efforts.

				Autocalibration inputs		
Parameter	***Value***	***Units***	***Symbol***	***Auto-cal***	***Min value***	***Max value***
Stoichiometry:						
Carbon	40	gC	gC	No	30	60
Nitrogen	7.2	gN	gN	No	5	9
Phosphorus	1	gP	gP	No	0.5	2
Dry weight	100	gD	gD	No	100	100
Chlorophyll	1	gA	gA	No	0.5	2
Inorganic suspended solids:						
Settling velocity	Manual	m/d	v_i	No	0.2	2
Oxygen:						
Reaeration model	Manual Determination of Appropriate Formula			No		
Temp correction	1.024		q_a			
Reaeration wind effect	None					
O2 for carbon oxidation	2.69	gO_2/gC	r_{oc}			
O2 for NH4 nitrification	4.57	gO_2/gN	r_{on}			
Oxygen inhib model CBOD oxidation	Exponential					
Oxygen inhib parameter CBOD oxidation	0.60	L/mgO2	K_{socf}	No	0.60	0.60
Oxygen inhib model nitrification	Exponential					
Oxygen inhib parameter nitrification	0.60	L/mgO2	K_{sona}	No	0.60	0.60
Oxygen enhance model denitrification	Exponential					
Oxygen enhance parameter denitrification	0.60	L/mgO2	K_{sodn}	No	0.60	0.60
Oxygen inhib model phyto resp	Exponential					
Oxygen inhib parameter phyto resp	0.60	L/mgO2	K_{sop}	No	0.60	0.60
Oxygen enhance model bot alg resp	Exponential					
Oxygen enhance parameter bot alg resp	0.60	L/mgO2	K_{sob}	No	0.60	0.60
Slow CBOD:						
Hydrolysis rate	0	/d	k_{hc}	No	0.05	0.25
Temp correction	1.047		q_{hc}	No	1	1.07
Oxidation rate	0.103	/d	k_{dcs}	No	0.05	0.25
Temp correction	1.047		q_{dcs}	No	1	1.07
Fast CBOD:						
Oxidation rate	10	/d	k_{dc}	No	0	10
Temp correction	1.047		q_{dc}	No	1	1.07
Organic N:						
Hydrolysis		/d	k_{hn}	Yes	0.05	0.3
Temp correction	1.07		q_{hn}	No	1	1.07
Settling velocity		m/d	v_{on}	Yes	0.05	0.25
Ammonium:						
Nitrification		/d	k_{na}	Yes	0.05	4
Temp correction	1.07		q_{na}	No	1	1.07
Nitrate:						
Denitrification		/d	k_{dn}	Yes	0.05	2
Temp correction	1.07		q_{dn}	No	1	1.07
Sed denitrification transfer coeff		m/d	v_{di}	Yes	0	1
Temp correction	1.07		q_{di}	No	1	1.07
Organic P:						
Hydrolysis		/d	k_{hp}	Yes	0.05	0.3
Temp correction	1.07		q_{hp}	No	1	1.07
Settling velocity		m/d	v_{op}	Yes	0.05	0.25

				Autocalibration inputs		
Parameter	***Value***	***Units***	***Symbol***	***Auto-cal***	***Min value***	***Max value***
Inorganic P:						
Settling velocity		m/d	v_{ip}	Yes	0	2
Sed P oxygen attenuation half sat constant		mgO_2/L	k_{spi}	Yes	0	2
Phytoplankton:						
Max Growth rate		/d	k_{gp}	Yes	1.5	3
Temp correction	1.07		q_{gp}	No	1	1.07
Respiration rate		/d	k_{rp}	Yes	0.05	0.5
Temp correction	1.07		q_{rp}	No	1	1.07
Death rate		/d	k_{dp}	Yes	0	1
Temp correction	1		q_{dp}	No	1	1.07
Nitrogen half sat constant	15	ugN/L	k_{sPp}	No	10	25
Phosphorus half sat constant	2	ugP/L	k_{sNp}	No	1	5
Inorganic carbon half sat constant	1.30E-05	moles/L	k_{sCp}	No	1.30E-06	1.30E-04
Phytoplankton use HCO3- as substrate	Yes					
Light model	Smith					
Light constant	57.6	langleys/d	K_{Lp}	No	40	110
Ammonia preference	15	ugN/L	k_{hnxp}	No	15	30
Settling velocity		m/d	v_a	Yes	0.05	0.5
Bottom Plants:						
Growth model	Zero-order					
Max Growth rate		$gD/m^2/d$ or /d	C_{gb}	Yes	1.5	200
Temp correction	1.07		q_{gb}	No	1	1.07
First-order model carrying capacity	100	gD/m^2	$a_{b,max}$	No	50	200
Basal respiration rate		/d	k_{r1b}	Yes	0.02	0.2
Photo-respiration rate parameter	0.39	unitless	k_{r2b}	No	0	0.6
Temp correction	1.07		q_{rb}	No	1	1.07
Excretion rate		/d	k_{eb}	Yes	0	0.5
Temp correction	1.07		q_{db}	No	1	1.07
Death rate		/d	k_{db}	Yes	0	5
Temp correction	1.07		q_{db}	No	1	1.07
External nitrogen half sat constant		ugN/L	k_{sPb}	Yes	100	500
External phosphorus half sat constant		ugP/L	k_{sNb}	Yes	25	100
Inorganic carbon half sat constant		moles/L	k_{sCb}	Yes	1.30E-06	1.30E-04
Bottom algae use HCO3- as substrate	Yes					
Light model	Half saturation					
Light constant		langleys/d	K_{Lb}	Yes	40	100
Ammonia preference		ugN/L	k_{hnxb}	Yes	15	30
Subsistence quota for nitrogen		mgN/gD	q_{0N}	Yes	0.36	1.44
Subsistence quota for phosphorus		mgP/gD	q_{0P}	Yes	0.05	0.2
Maximum uptake rate for nitrogen		mgN/gD/d	r_{mN}	Yes	350	1500
Maximum uptake rate for phosphorus		mgP/gD/d	r_{mP}	Yes	50	200
Internal nitrogen half sat ratio			$K_{qN,ratio}$	Yes	1.05	5
Internal phosphorus half sat ratio			$K_{qP,ratio}$	Yes	1.05	5
Nitrogen uptake water column fraction	1		$N_{UpWCfrac}$	No	0	1
Phosphorus uptake water column fraction	1		$P_{UpWCfrac}$	No	0	1
Detritus (POM):						
Dissolution rate		/d	k_{dt}	Yes	0.05	5
Temp correction	1.07		q_{dt}	No	1.07	1.07
Settling velocity	0.4033805	m/d	v_{dt}	Yes	0.05	0.5

Table 11.10. Weighting factors for each constituent used to calculate the fitness in model calibration.

Parameter	Weighting Factor
DO (mgO_2/L)	5
CBODs (mgO_2/L)	1
Norg (ugN/L)	2
NH4 (ugN/L)	3
NO3 (ugN/L)	3
Porg (ugN/L)	2
Inorg P (ugP/L)	4
Phyto (ugA/L)	1
Alk ($mgCaCO_3$/L)	4
pH	4
TN (ugN/L)	3
TP (ugP/L)	3
TSS (mgD/L)	1
CBODu (mgO_2/L)	1
DO (mgO_2/L) - Min	5
DO (mgO_2/L) - Max	5
CH-A - Min	1
CH-A - Max	1

MODEL VALIDATION/CORROBORATION

At two locations (Silver Creek and Fairview) validation or corroboration data sets (identical to the calibration data sets) were collected during a different time period. The objectives of these data sets were to determine if the model calibrations held during a different time period under somewhat different conditions. For model corroboration we updated the boundary condition, point inflow, and weather data to coincide with the conditions during the validation time period. All other site specific information (e.g., channel characteristics) and parameters set during calibration were held constant. The exception was SOD which can change during the year due to the transfer of oxygen demanding material into the study reach. The SOD value for the validation period was again estimated based on ER- GPP at station E.

Findings/Recommendations/Suggested Future Work

In general, we have found that the models resulting from this study have been able to meet the diverse intended uses. For example, a number of the models have already been foundational in developing WLAs and they are currently in the process of being used to assist in statewide nutrient criteria development. However, given the generic nature of the data collection and automatic calibration methods necessary to meet these varied needs and applications, there is at times significant uncertainty in important mechanisms and therefore in predictions. In some circumstances, additional data collection efforts, sensitivity and uncertainty analyses will be necessary to ensure the appropriate confidence in model predictions. Recommendations and suggested future efforts have also been identified.

Data Collection

In general, one of the most important lessons learned from this effort was the need for a larger number of samples due to the short timescale of the data collection campaigns (2-3 days). These data collection efforts focused on collecting data during presumed steady state conditions which further led to the assumption that many of the data types we gathered would not vary significantly throughout each day (including flow and water quality). The exceptions from this assumption were temperature, DO, pH, specific conductance, and chlorophyll a which were measured at small time increments over 3-4 days in an effort to get a good understanding of this daily variability. While the streams themselves and the conditions at station A and B were relatively stable during these late summer time periods, the conditions downstream of many of the WRFs were not. Based on these studies, stable conditions do not exist for many of these plants and the loads are highly variable throughout the day. This becomes critical to consider when sampling and modeling effluent dominated systems (e.g., Silver Creek, Moroni). This variability caused significant problems during model population and calibration due to samples often not representing the average conditions and resulting in very different values between days where grab samples were collected. The Washington Department of Ecology generally samples twice a day for two days in a row in the stream. They also use a 24-hour composite for two days from the WRF effluent. Further, they average three benthic algae samples at randomly selected sites where periphyton are present. A similar approach may be warranted within the state of Utah, however, the representativeness of this sampling regime should be investigated.

We also found that samples taken at different locations often did not coincide with the samples taken at a calibration location. To illustrate some of the disconnects we encountered, assume a sample is taken at station E (calibration location) at 10:00 am and this corresponded to the WRF release at 8 am (i.e., there is a 2 hour travel time between the WRF and Station E). The sample then taken at the WRF for the modeling occurs at 10:30 am. This and the measured flow value from the WRF is then used to calculate the load within the model that gets decayed and transported downstream to station E (the calibration location). In this example, if the WRF effluent varies significantly over short time periods you can see how the samples used in model forcing and calibration can easily be disconnected and influences model calibration and ultimately interpretation.

These sorts of issues could be dealt with by taking more samples throughout the study period (e.g., 3 per day) which would provide a much better understanding of the mean and variance throughout the entire study period. However, the constituents requiring higher frequency sampling will likely be site specific and depend on the loads impacting the system (e.g., highly variable WRF versus a lagoon system). Due to the WRF variability, the 2 data points gathered often times provided a large range of possible concentrations and made it difficult to decide on an appropriate representative average to be used in the load estimates or in calibration. Another concern was that there were many times that data points were either missing, resulted in non-detects, etc. These missing data left us with even less information for model population and/or calibration. In the future when dealing with effluent dominated streams with highly variable loads, it would be more appropriate to gather time-variable data and use the newly developed version of QUAL2Kw that allows for non-steady flow with a continuous simulation option over a 365 day simulation period.

When it came to understanding loads, we also identified the need to collect enough flow information to ensure an appropriate water balance throughout the study reach. Since the ability to predict accurate concentrations hinges on correct volumes, in cases where the inflows were variable or discharge measurements showed variability, more measurements were necessary. This includes ensuring accurate flow estimates at each of the study sites throughout the study period and may require the use of various flow measurement methods (e.g., slug injections rather than velocity area methods). Good flow data provides information regarding the appropriateness of the steady flow assumption and also provides a more solid estimate of average flow conditions if there is variability present. This again highlights the need to understand the variability in WRF effluent. To assist in these efforts and all load allocation decision making (e.g., TMDLs or NPDES permits), we recommend that the state requires WRFs to track subhourly effluent rates and provide these to the state quarterly. It may be worthwhile to also have them install a water quality sonde and track the effluent DO, temperature, specific conductance, and pH since these data provide information regarding the plant effluent concentration variability and potential plant upsets.

The other key concern identified within these data had to do with analytical methods and the associated errors. For example, the sCBOD method detection limits are 3 or 5 mg/L depending on the lab. With low sCBOD both in WRF effluent and many of the streams not being highly influenced by high BOD loads, we were forced to calculate the actual BOD loads (from the WRF or at the upstream boundary condition) based on concentrations assumed to be half the method detection limit. While other, less biased techniques exist to handle these censored values (i.e., trimmed mean, Winsorized mean, Cohen's maximum likelihood method), they are incapable of handling cases where more than 25% of the data is censored. These types of assumptions lead to significant errors in loads and the resulting sCBOD predictions. This may become a significant enough issue that new analytical methods need to be developed. Similarly, we ran into issues with analytical error when it came to estimating some constituent concentrations based on differences (e.g., Organic N, Organic P, and detritus). The various sources of sampling and analytical error can produce significant errors in model loads and model calibration. This was particularly important given the limited number of samples and again illustrates the need for additional sampling throughout the study period. Further, we identified that a measure of VSS should be included in the sampling protocol to better estimate detritus concentrations. Detritus could then become part of the fitness statistic and used in calibration.

Model Population/Calibration

Many of the issues associated with data collection have an obvious link to the success of model calibration, the ability to minimize model uncertainty, and the utility in decision making. Other concerns were identified that were more specifically related to model population or calibration.

A key concern was the very short spatial scale over which data were collected. In an effort to minimize the influence of tributaries, withdrawals, etc. and to meet the needs associated with quantifying open water metabolism, data were collected over short reaches based on Eqn. 7. This equation provides an estimated reach length where half of the oxygen has exchanged with the atmosphere via reaeration [*Grace and Imberger*, 2006]. While these distances were appropriate for the metabolism estimates, the associated short travel times resulted in many of the chemical reactions having minimal influence within the study reach. In other words, over these travel time scales many reactions had minimal impact on instream concentrations resulting in relatively insensitive parameters. To address this concern, additional data sets were gathered in summer 2011 and the reach lengths were extended as much as possible given the tributary inflows, diversions, etc. that would require even more extensive data collection. For most future modeling applications (with the exception of WLA analyses) it would be best to make reach lengths as long as possible for the modeling study and deploy dissolved oxygen sensors within the reach at the optimal lengths based on Eqn. 7. If the approach provided by Grace and Imberger [2006] is still used, we suggest maximizing the multiplier (use 3 instead of 0.693, increasing importance of instream processes from 50% to 95%, $-\ln(.05) \approx 3$) to ensure longer study reaches for the modeling and maintain the ability to still use the one or two station metabolism methods.

Another key issue identified in QUAL2Kw modeling is the need to decrease the number of parameters that are autocalibrated. As discussed previously, when possible, parameters should be measured or estimated for the study site of interest. Some key rates that can be estimated include:

1. BOD decomposition rates (k_d) could be estimated for each of the WRF types (lagoons, oxidation ditch, membrane) in Utah.
2. Nitrification rates could be estimated for each study site.
3. Photosynthetic Active Radiation (PAR) attenuation within the water column given the importance of bottom algae in many of these systems.

The other key parameter, at least in some systems, is SOD. While we established a method of estimating SOD using dissolved oxygen measurements and metabolism methods, further investigation into the assumptions made regarding the minimal influence of other oxygen demanding reactions that are reflected by the in-situ dissolved oxygen measurements and the amount of autotrophic respiration should be considered. Further, the application of these methods to all systems needs to be investigated.

When it comes to autocalibration, there is an obvious need to decrease the number of parameters and potentially come up with narrower ranges to confine autocalibration estimates. In these model applications, over 30 parameters are being optimized. This number is extremely high, but without more information regarding which parameters are unimportant, it is not clear which should be dropped from the autocalibration. We have seen that the phytoplankton parameters seem to be insensitive. However, the bottom plants predictions are very important in many Utah streams and it is unknown which bottom algae parameters are sensitive and should be included in autocalibration. Approximately 15 of the parameters being optimized are associated with bottom plant growth and there is minimal information regarding the spatial and temporal concentrations to be used in calibration. Benthic algae carbon, nitrogen, and phosphorus ratios were established within some streams to provide an understanding of autotrophic nutrient limitation and provide insight into the heterotrophic resource quality. These data could be useful in bottom algae parameter estimation, however, they show there is significant spatial and temporal variability in stoichiometry along study reaches. This presents additional challenges in developing the appropriate sampling approaches to collect representative data at the reach or sub reach scales. The utility of these data types and sampling techniques need to be further investigated. Additionally, the number of simulation days influence the bottom plant concentrations and guidance regarding how best to set the simulation time period should be developed.

We recommend a sensitivity analysis be completed for all these case studies to determine if there can be global reduction (meaning for all study sites) in the number of parameters autocalibrated and then which parameters are insensitive in different systems. This will result in a reduction of parameter sets that produce similar output responses. Within this effort, it would be important to identify which output parameters are important and influence the fitness statistic since the objective function (i.e., fitness) guides the calibration. If output values are not sensitive, they can influence the calibration algorithm performance.

In these applications, given the number of calibration parameters, short travel times, and limited amount of data, the resulting calibrations may or may not be appropriate for different circumstances. We say this because there were some consistent findings that suggest that we are missing key processes, some parameters included in calibration were insensitive, or our approach to autocalibration may need some refinement. For example, the autocalibration algorithm consistently set sediment denitrification rates and inorganic phosphorus settling rates to relatively high values (Table 11. 11). Both of these parameters basically provide a way to remove N and P from the water column, but in general this is done in way that does not provide any insight into underlying mechanisms. In other words, these model terms are merely a N and P sink. It is recommended that the influence of these parameters in autocalibration be investigated. If these additional N and P sinks truly exist, there is a need to investigate which mechanisms are not present within the model but are being consistently observed in these systems. Another interesting result of these calibrations are predicted pH values that are consistently too high. This can be important in the ability to predict other constituent concentrations and the mechanisms leading to this should be revisited.

Table 11.11. Range of parameters found within the 9 QUAL2Kw models within Utah.

	Brigham City	*Fairview*	*Moroni*	*Oakley City*	*Price*	*Silver Creek*	*Spanish Fork*	*Tremonton*	*Wellsville*			
Fitness	4.2	3.9	3.9	6.0	6.3	12.0	3.3	6.0	9.9			
Parameter										***Overall Results***		
										Average	***Min***	***Max***
Inorganic suspended solids:												
Settling velocity	0.001	2	1.5	0.001	0.2	2	0.2	0.2	0.2	0.700	0.001	2.000
Oxygen:												
Reaeration model	Internal	Tsivoglou-Neal	Owens-Gibbs	Internal	USGS(pool-riffle)	Tsivoglou-Neal	USGS(channel-control)	Internal	Owens-Gibbs			
Slow CBOD:												
Oxidation rate	0.23	0.10	0.10	0.19	0.10	0.10	0.10	0.10	0.10	0.13	0.10	0.23
Organic N:												
Hydrolysis	0.08	0.30	0.26	0.22	0.25	0.08	0.25	0.27	0.09	0.20	0.08	0.30
Settling velocity	0.13	0.23	0.21	0.19	0.19	0.11	0.07	0.22	0.15	0.17	0.07	0.23
Ammonium:												
Nitrification	3.60	4.00	1.74	3.48	0.05	3.10	3.84	0.93	0.87	2.40	0.05	4.00
Nitrate:												
Denitrification	1.94	1.06	1.44	0.10	0.31	0.89	0.44	1.01	0.89	0.90	0.10	1.94
Sed denitrification transfer coeff	0.32	0.04	0.98	0.15	0.74	0.99	0.89	0.03	0.56	0.52	0.03	0.99
Organic P:												
Hydrolysis	0.10	0.09	0.08	0.15	0.13	0.12	0.11	0.28	0.24	0.15	0.08	0.28
Settling velocity	0.23	0.05	0.20	0.11	0.13	0.08	0.15	0.10	0.11	0.13	0.05	0.23
Inorganic P:												
Settling velocity	0.07	1.26	1.97	1.90	1.95	1.82	1.50	0.09	0.61	1.24	0.07	1.97
Sed P oxygen attenuation half sat constant	1.21	1.41	0.47	0.62	0.10	1.35	1.23	2.00	1.56	1.11	0.10	2.00
Phytoplankton:												
Max Growth rate	2.2	2.0	2.8	1.8	2.4	2.9	2.8	1.8	2.9	2.4	1.8	2.9
Respiration rate	0.2	0.3	0.2	0.1	0.2	0.4	0.2	0.1	0.1	0.2	0.1	0.4
Death rate	0.4	0.6	0.2	0.7	0.1	0.7	0.8	0.0	0.1	0.4	0.0	0.8
Ammonia preference	15.0	25.3	26.2	15.0	16.8	19.7	16.2	23.2	19.9	19.7	15.0	26.2
Settling velocity	0.1	0.4	0.4	0.4	0.1	0.3	0.2	0.1	0.1	0.2	0.1	0.4
Bottom Plants:												

	Brigham City	***Fairview***	***Moroni***	***Oakley City***	***Price***	***Silver Creek***	***Spanish Fork***	***Tremonton***	***Wellsville***			
Max Growth rate	10.2	39.5	15.7	85.0	15.8	60.6	39.2	161.1	8.6	48.4	8.6	161.1
Basal respiration rate	0.15	0.19	0.11	0.07	0.07	0.20	0.20	0.05	0.09	0.13	0.054	0.20
Photo-respiration rate parameter	0.39	0.01	0.01	0.39	0.01	0.01	0.01	0.01	0.01	0.09	0.010	0.39
Excretion rate	0.25	0.11	0.28	0.12	0.33	0.07	0.00	0.36	0.39	0.21	0.003	0.39
Death rate	0.65	0.07	2.64	0.01	1.67	0.01	0.01	4.46	4.03	1.50	0.005	4.46
External nitrogen half sat constant	389	253	374	264	350	180	465	320	184	309	180	465
External phosphorus half sat constant	47	68	48	63	67	76	56	57	90	64	47	90
Inorganic carbon half sat constant	1.7E-05	9.1E-05	1.2E-04	1.1E-04	7.4E-05	3.4E-05	7.8E-05	9.0E-05	2.5E-05	0.000	0.000	0.000
Light model	Half saturation	Smith	Smith	Half saturation	Smith	Smith	Smith	Smith	Smith			
Light constant	55	66	64	87	69	57	48	46	55	61	46	87
Ammonia preference	16	26	26	15	18	23	23	30	17	21	15	30
Subsistence quota for nitrogen	1.4	0.7	0.9	1.0	0.9	1.0	0.8	0.7	1.4	1.0	0.7	1.4
Subsistence quota for phosphorus	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.2
Maximum uptake rate for nitrogen	481	431	427	1405	744	764	957	724	1056	776	427	1405
Maximum uptake rate for phosphorus	117	101	175	184	145	163	98	124	146	139	98	184
Internal nitrogen half sat ratio	1.8	1.2	1.6	4.4	1.6	4.4	3.5	1.5	2.1	2.5	1.2	4.4
Internal phosphorus half sat ratio	2.1	3.5	1.3	4.8	5.0	3.2	3.9	1.4	2.8	3.1	1.3	5.0
Detritus (POM):												
Dissolution rate	3.70	1.58	4.75	1.63	0.28	4.68	1.07	0.07	1.66	2.16	0.07	4.75
Temp correction	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07			
Settling velocity	0.07	0.42	0.22	0.16	0.07	0.16	0.49	0.11	0.48	0.24	0.07	0.49
User-defined autocalibration parameters (optional)												
Prescribed SOD (gO2/m2/day)	0.0	0.0	17	0	0.1	11	0	0	5	3.7	0	17

Use of Models in Support of Nutrient Criteria Development

As mentioned previously, this work was part of a greater effort to provide information that guides the development of nutrient criteria for the state of Utah. The goal is to evaluate changes in ecosystem structure (fish and macroinvertebarte communities), ecosystem function (whole stream metabolism, nutrient limitation, organic matter storage and decompositions rates), and water chemistry and quality above and below each of the treatment plant discharges. The proposed numeric nutrient criteria will consist of nitrogen and phosphorus limits, as well as other response indicators of primary production, ecosystem composition, and ecosystem function (Figure 11.4).

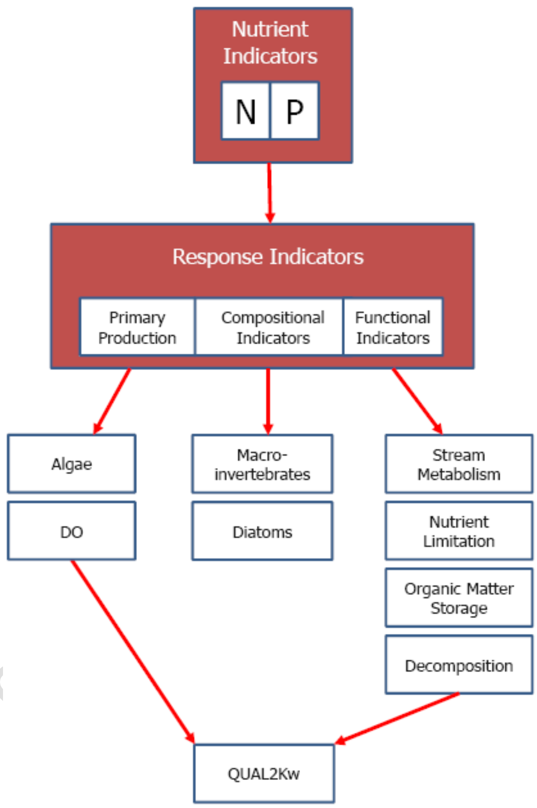


Figure 11.4. Numeric indicators of excess nitrogen and phosphorus pollution

Using the QUAL2Kw models built and calibrated for each study site it is possible to predict the effects of nutrient addition, or removal, on these response variables. The model provides an additional line of evidence for the development of numeric nutrient criteria by linking excess nitrogen and phosphorus levels in streams to thresholds in response variables such as algal growth and DO.

The QUAL2Kw models will be applied to nutrient criteria development using critical conditions for flow, meteorology, and water quality, either the same as the calibration conditions or generated similar to those for wasteload analyses (UDEQ, 2012). Within the model, inorganic nitrogen and phosphorus concentrations will be adjusted to identify the concentration that will result in just meeting the threshold level for each response indicator. Each nutrient will be analyzed separately (i.e., when conducting the phosphorus criteria analysis, nitrogen concentration will be set high enough so as not to limit algal growth).

The Following are potential linkages between QUAL2Kw output and each response indicator (Figure 11.4):

1. Primary Productivity
   a. Benthic Algae (as measured either by Chlorophyll a or ash free dry mass [AFDM]): QUAL2Kw direct output, expressed as either Chlorophyll a or total algal biomass, will be compared to the recreation based threshold of 150 mg/m^2 of Chl a.
   b. Dissolved Oxygen: QUAL2Kw direct output will be compared to existing water quality standards for DO with and without early life stages present.
   c. pH: QUAL2Kw direct output versus maximum pH and diel change in pH.
2. Compositional Indicators: QUAL2Kw does not currently address compositional indicators.
3. Functional Indicators
   a. Stream Metabolism: QUAL2Kw direct output of gross primary productivity (GPP), expressed as gO_2/m^2, will be compared to thresholds developed by the ecological study.
   b. Nutrient Limitation: Determine whether nitrogen or phosphorus is the limiting nutrient at critical condition.
   c. Organic Matter Storage: Total organic matter storage in the sediments is not currently a standard output of QUAL2Kw. Typically, much of the organic matter gets deposited in the sediments outside of the simulation period and is added as prescribed SOD in the model.
   d. Decomposition Rate: Decomposition rate of organic matter does not currently vary by nutrient concentration in QUAL2Kw. Once the scientific literature quantifies the relationship between decomposition rate (both in the water column and sediments), will inquire with Washington DOE whether this functionality can be incorporated into the model.

The study output is anticipated to be nitrogen and phosphorus criteria that meet the proposed response indicator thresholds for algal biomass, dissolved oxygen, pH and gross primary productivity. By applying this to multiple models in different physiographic settings, we will have a range of nitrogen and phosphorus criteria for stream and river systems in Utah.

LITERATURE CITED

Acuna, V., Giorgi, A., Munoz, I., Uehlinger, U. and Sabater, S. 2004. Flow extremes and benthic organic matter shape the metabolism of a headwater Mediterranean stream. *Freshwater Biology*. 49(7):960-971.

Allan, J.D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology Evolution and Systematics* 35, 257–284.

Allegeier, J.E., Rosemond, A.D., and Layman, C.A. 2011. The frequency and magnitude of non-additive Responses to multiple nutrient enrichment. *Journal of Applied Ecology*. 48:821-834.

Arbia, G. and G. Lafratta. 1997. Evaluating and updating the sample design in repeated environmental surveys: monitoring air quality in Padua. *Journal of Agricultural, Biological, and Environmental Statistics* 2:451-466.

Aristegi, L., O. Izagirre, and A. Elosegi (2009), Comparison of several methods to calculate reaeration in streams, and their effects of estimation of metabolism., *Hydrobiologia*, *635*, 113-124.

Ascough, J. C., H. R. Maier, J. K. Ravalico, and M. W. Strudley. 2008. Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modeling* 219:383-399

Baker, M.E. & King, R.S. 2010. A new method for detecting and interpreting biodiversity and ecological thresholds. *Methods in Ecology and Evolution*. 1:25-37.

Barbour, M.T., Swietlik, W.F. Jackson, S.K. Courtemanch, D.L. Davies, S.P. & Yoder, C.O. 2000 Measuring the attainment of biological integrity in the USA: A critical element of ecological integrity. Hydrobiologia. 422:653-664

Barbour, M.T., Gerritsen, J., Snyder, B.D. and Stribling, J.B. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water; Washington D.C.

Barko, J.W. and Smart. 1981. Interelationship between the growth of *Hydrilla verticillata* (L.f.) Royale and sediment nutrient availability. *Aquatic Botany*. 32:205-216.

Barton, D. R. and S. M. Smith. 1984. Insects of extremely small and extremely large aquatic habitats. Pages 456-483 *in* V. H. Resh and D. M. Rosenberg, editors. The Ecology of Aquatic Insects. Praeger Publishing, New York.

Bernhardt, E.S. and Likens, G.E. 2004. Controls on periphyton biomass in heterotrophic streams. *Freshwater Biology* 49:14-27

Benstead, J.P., Rosemond, A.D., Cross, W.F., Wallace, J.B., Eggert, S.L., Suberkropp, K., Gulis, V., Greenwood, J.L. and Tant, C.J. 2009. Nutrient enrichment alters storage and fluxes of detritus in a headwater stream ecosystem. Ecology. 90(9):2556-2566.

Bernot, M. J., D. J. Sobota, R. O. Hall, P. J. Mulholland, W. K. Haapala, J. R. Webster, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, N. B. Grimm, S. K. Hamilton, S. L. Johnson, W. H. McDowell, J. L. Meyer,B. Peterson, G. C. Poole, H. M. Valett, J. R. Webster, C. Arango, J. J. Beaulieu, A. J. Burgin, C. L. Crenshaw, A. M. Helton, L. Johnson, J. Merriam, B. R. Nitankederlehner, J. M. O'Brien, J. D. Potter, R. W. Scheibley, and K. Wilson. 2010. Inter-regional comparison of land-use effects on stream metabolism. Freshwater Biology 55: 1874-1890.

Biggs, B. J. F. 2000. Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. *Journal of North American Benthological Society* 19:17–31.

Bilby, R.E. and Likens, G.E. 1980. Importance of organic debris dams in the structure and function of stream ecosystems. *Ecology*. 61(5):1107-1113.

Bischoff, J., P. Massaro, and J. Strom (2010), Jewitts Creek dissolved oxygen TMDL, edited by C. R. O. o. Water, Wenck Associates, Inc.

Bonin, H.L, Griffiths, R.P. and Caldwell, B.A. 2000. Nutrient and microbiological characteristics of fine benthic organic matter in mountain streams. *Journal of the North American Benthological Society*. 19(2):235-249.

Bott, T.L., Montgomery D.S., Newbold, J.D., Arscott, D.B., Dow, C.L., Aufdenkampe, A.K., Jackson, J.K. and Kaplan, L.A. 2006. Ecosystem metabolism in streams of the Catskill mountains (Delaware and Hudson River watersheds) and Lower Hudson Valley. *Journal of the North American Benthological Society*. 25(4):1018-1044.

Bowie, G. L., W. B. Mills, D. B. Porcella, C. L. Campbell, J. R. Pagenkopf, G. L. Rupp, K. M. Johnson, P. W. H. Chan, S. A. Gherini, and C. E. Chamberlin (1985), Rates, Constants, and Kinetic Formulations in Surface Water Quality Modeling edited by E. R. Laboratory, United States Environmental Protection Agency, Athens, GA.

Box, G. E. P. and N. R. Draper. 1987. Empirical Model-Building and Response Surfaces. *in* Wiley Series in Probability and Statistics. Wiley, New York, 688 pp.

Breiman, L. 2001. Random Forests. *Machine Learning*. 45(1):5-32.

Breiman, L. 1996. Bagging predictors. *Machine Learning*. 24(2):123-140.

Brooks A.J., B.C. Chessman and T. Haeusler 2011 .Macroinvertebrate traits distinguish unregulated rivers subject to water abstraction. *Journal of the North American Benthological Society*, 30, 419–435.

Bowden, W.B., J.M. Glime and T. Riis. 2006. Macrophytes and Bryophytes. Pages 381-414 in F.R. Hauer and G.A.Lamberti (eds.). Methods in Stream Ecology. Elsevier/Academic Press, San Diego CA. Chambers, P.A., Prepas, E.E., Hamilton, H.R. and Bothwell. 1991. Current Velocity and its effect on aquatic macrophytes in flowing waters. *Ecological Applications*. 1(3):249-257.

Burdon, F. J., A. R. McIntosh, and J. S. Harding. 2013. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. *Ecological Applications* 23:1036–1047.

(Capps et al. 2011)

Carr, J.F. 1962. Dissolved oxygen in Lake Erie, past and present. Publ. Great Lakes Res. Div., Univ. Mich., 9:1-14.

(Carr et al. 2005).

(Chapra, S. C. (1997), *Surface Water Quality Modeling*, McGraw-Hill, Boston, MA.

(Chatelot et al. 1999)

Chao, C. T. and S. K. Thompson. 2001. Optimal adaptive selection of sampling sites. *Environmetrics* 12:517-538.

CH2M Hill Inc. 2012. Study of the Statewide Benefits of Nutrient Reduction. Utah Department of Environmental Quality, Division of Water Quality.

Cho, J. H., and S. R. Ha (2010), Parameter optimization of the QUAL2K model for a multiple-reach river using an influence coefficient algorithm., *Science of The Total Environment*, *408*(8), 1985-1991.

(Clarke at al. 2008).

(Clark et al. 2000).

Connelly, N.M., Crossland, M.R. and Pearson, R.G. 2004. Effect of low dissolved oxygen on survival, emergence, and drift of tropical stream macroinvertebrates. *Journal of the North American Benthological Society*. 23(2):251-270.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Publications, Agencies and Staff of the U. S. Department of Commerce*. Paper 189.

Cross, W.F. Wallace, J.B., Rosemond, A.D. and Eggert, S.L. 2006. Whole-system nutrient enrichment increase secondary production in a detritus-based ecosystem. *Ecology*. 87:1556-1565.

(Cuffney et al. 2006)

Cuffney, T.F., Zappia, H., Giddings, E.M.P. & Coles, J.F. 2005. Effects of urbanization on benthic macroinvertebraate assemblages in contrasting environmental settings: Boston, Massachusetts, Brimingham, Alabama and Salt Lake City, Utah. *American Fisheries Symposium*, 2005:361-407.

Cuffney, T.F., Gurtz, M.E. & Meador, M.R. 1993. Methods for collecting benthic invertebrate samples as part of the Nation Water-Quality Assessment program: U.S. Geological Survey Open-File Report 93-406, 66p.

Cullen, A. C. and H. C. Frey. 1999. Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. Springer Science and Business Media. 335 pp..

Davie, S.P. and Jackson, S.K. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications*. 16(4):1251-66.

Dambacher, J. M., H. W. Li, and P. A. Rossignol. 2002. Relevance of community structure in assessing indeterminacy of ecological predictions. *Ecology* 83:1372-1385.

Dewson, Z.S., A.B.W James. And R.G. Death. 2007. A review of the consequences of decreased flow for instream habitat and macroinvertebrates. *Journal of the North American Benthological Society*, 26, 401–415.

Dodds, W. K. 2007. Trophic state, eutrophication and nutrient criteria in streams. *Trends in Ecology and Evolution* 22(12):669-676.

Dodds, W.K., Smith, V.H. & Lohman, K. 2002. Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. *Canadian Journal of Fisheries and Aquatic Sciences*. 59:865-874.

Dodds, W.K. & Welch, E.B. 2000. Establishing nutrient criteria in streams. *Journal of the North American Benthological Society*. 19(1):186-196.

Downes, B. J. 2010. Back to the future: little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems. *Freshwater Biology*. 55: 60-79.

(Dudley et al. 2000)

Edmondson, W. T., G. C. Anderson, and D. R. Petersen. 1956. Artificial eutrophication of Lake Washington. *Limnology and Oceanography* 1:4753.

Elser, J.J., Marzolf, E. and Goldman, C.R. 1990. The roles of phosphorus and nitrogen in limiting phytoplankton growth in freshwaters: a review of experimental techniques. *Canadian Journal of Fisheries and Aquatic Sciences* 47:1468-1477.

Environmental Protection Division (1989), The Amplified Long-Term BOD Test, Protocol/Procedures and Test Specifications. , edited by D. o. N. Resources, p. 38 pp., Atlanta, GA.

Evans, G & Jones, L. 2001. *Economic impact of the 2000 red tide on Galveston County, Texas: A Case Study*. College Station, TX: Department of Agricultural Economics, Texas A&M University.

Fairchild, G.W., Lowe, R.L. and Richardson, W.B. 1985. Algal periphyton growth on nutrient diffusing Substrates: an in situ bioassay. *Ecology* 66(2)465-472.

Fausch, K.D., Lyons, J., Karr, J.R. & Angermeier, P.L. 1990. Fish Communities as indicators of environmental degradation. *American Fisheries Society Symposium.* 8:123-144.

Federov, V. V. and C. Nachtsheim. 1995. Optimal designs for time-dependent responses. Pages 3-13 *in* C. P. Kitsos and W. G. Muller, editors. Proceedings of MODA4. Physica-Verlag, Heidelberg, Germany.

(Fereira et al. 2006).

Ferson, S. 2002. RAMAS Risk Calc: Risk Assessment with Uncertain Numbers. Lewis Press, Boca Raton, FL.

Ferson, S., W. Root, and R. Kuhn. 1999. RAMAS Risk Calc: Risk Assessment with Uncertain Numbers. Applied Biomathematics, Setauket, NY.

Ferson, S. 1996. What Monte Carlo methods cannot do? *Human and Ecological Risk Assessment* 2:990-1007.

Ferson, S. and L. R. Ginzburg. 1996. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety* 54:133-144.

Findlay, S.E.G., Sinsabaugh, R.L., Sobczack, W.V. and Hoostal, M. 2003. Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter. *Limnology and Oceanography.* 48(4):1608-1617.

Fisher, S.G. and Likens, G.E. 1973. Energy flow in Bear Brook, New Hampshire: an integrative approach to stream ecosystem metabolism. *Ecological Monographs.* 43(4):421-439.

Flynn, K and Suplee, M.N. 2011. Using a computer water quality model to derive numeric nutrient criteria: Lower Yellowstone River. WQPBDMSTECH-22. Helena, MT: Montana Department of Environmental Quality.

Folt, C. L., C. Y. Chen, M. V. Moore, and J. Burnaford. 1999. Synergism and antagonism among multiple stressors. *Limnology and Oceanography* 44:864–877.

Florida Department of Environmental Protection. 2012. Technical support document: development of numeric nutrient criteria for Florida lakes, spring vents and streams.

Francouer, S.N. 2001. Meta-analysis of lotic nutrient amendment experiments: detecting and quantifying subtle responses. *Journal of the North American Benthological Society* 20(3):358-368.

(Freeman et al. 2007).

(Fritz et al. 2004),

Frey, H. C. 1993. Separating variability and uncertainty in exposure assessment: motivations and methods. Paper No. 93-79.01. Proceedings of the 86th Annual meeting, Air and Waste Management Association, Pittsburgh, PA.

Goodman, S. 2008. A dirty dozen: twelve p-value misconceptions. *Seminars in Hematology* 45:135-140.

Grace, M.G and Imberger, S. 2006. Stream metabolism: performing and interpreting measurements.

Grattan, R.M. and Suberkropp, K. 2001. Effects of nutrient enrichment on yellow poplar leaf decomposition and microbial activity in streams. *Journal of the North American Benthological Society*. 20:33-43.

(Greenwood and Rosemond 2011).

Gregory, R., L. Failing, M. Harstone, G. Long, T. McDaniels, and D. Ohlson. 2012. Structured Decision Making: a Practical Guide to Environmental Management Choices. Wiley-Blackwell, Hoboken, NJ, 312 pp.

Gulis, V. and Suberkropp, K. 2002. Leaf litter decomposition and microbial activity in nutrient-enriched and unaltered reaches of a headwater stream. *Freshwater Biology*. 48(1):123-134.

Hale, S.S. & Helthse, J.F. 2008. Signals from the benthos: development and evaluation of a benthic Index for the near shore gulf of Maine. *Ecological Indicators*. 8:338-350.

Hall, R.O. 2011. Stream metabolism workshop. Utah State University, Logan Utah. Unpublished data.

Hall, R.O., Wallace, J.B. and Eggert, S.L. 2000. Organic matter flow in stream food webs with reduced detrital resource base. *Ecology*. 81:3445-3463.

(Hall and Tank 2003)

Hall, R.O. and Meyer, J.L. 1998. The trophic significance of bacteria in a detritus-based stream food web. *Ecology*. 79(6):1995-2012.

Hansen, R. A., D. D. Hart, and R. A. Merz. 1991. Flow mediates predator-prey interactions between triclad flatworms and larval blackflies (Diptera: Simuliidae). *OIKOS* 46:88-92.

Harpole, W.S., Hgai, J.T., Seabloom, E.W., Borer, E.T., Bracken, M.E., Elser, J.J., Gruner, D.S., Hillenbrand, H., Shurin, J.B. and Smith, J.E. 2011. *Ecology Letters*. 14(9):963-62.

Hart, D. D. and C. M. Finelli. 1999. Physical-biological coupling in streams: the pervasive effects of flow on benthic organisms. *Annual Review of Ecology and Systematics* 30:363-395.

Hart, D. D., and Robinson, C.T. 1990. Resource limitation in a stream community: phosphorus enrichment effects on periphyton and grazers. Ecology 71:1494-1502.

Hawkins, C.P., Norris, R.H., Hogue, J.N. & Feminella, J.W. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*. 10 (5):1456-1477.

Hawkins, C.P., Olson, J.R. and Hill, R.A. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society*.29(1):312-343.

Hayes, K. R., H. M. Regan, M. A. Burgman, R. H. Devlin, A. R. Kapusciniski, and J. M. Dambacher. 2006. Methods to address uncertainty in ecological risk assessment of genetically modified organisms. http://www.gmo-safety.eu/pdf/biosafenet/Hayes.pdf.

Heathwaite A.L. 2010. Multiple stressors on water availability at global to catchment scales: understanding human impact on nutrient cycles to protect water quality and water availability in the long term. *Freshwater Biology 55*: 241–257.

Hecky, R.E. and Kilham, P. 1988. Nutrient limitation of phytoplankton in freshwater and marine Environments: A review of recent evidence on the effects of enrichment. *Limnology and Oceanography* 33(4):796-822.

(Helsel and Lee 2006).

Herlihy, A.T., Paulsen, S.G., Van Sickle, J.V., Stoddard, J.L., Hawkins, C.P. and Yuan, L.L. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approachat a continental scale. *Journal of the North American Benthological Society*. 27(4):860-877.

Hill, B.H., Hall, R.K., Husby, P., Herlihy, A.T. and Dunne, M. 2000. Interregional comparisons of sediment microbial respiration in streams. *Freshwater Biology*. 44(3): 213-222.

Hill, W.R. and Fanta, S.E. 2008. Phosphorus and light colimit periphyton growth at subsaturating Irradiances. *Freshwater Biology* 53(2):215-225.

Hill, W.R., S.E. Fanta, and B. J. Roberts. 2009. Quantifying phosphorus and light effects in stream algae. *Limnology and Oceanography* 54(1):368–380.

(Hoagland et al. 2006).

Hoagland, P., Anderson, D.M., Kaoru, Y. & White, A.W. 2002. The economic effects of harmful algal blooms in the United States, assessment issues, and information needs. *Estuaries*. 25(4b): 819-837.

Hobbs, R. J., J. E. Higgs, and J. A. Harris. 2006. Novel ecosystems: implications for conservation and restoration. *Trends in Ecology and Evolution* 24(11):559-605.

Holtgrieve, G.W., Schindler, D.E., Branch, T.A., A'mar, Z.A. (2010), Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics. Limnology and Oceanography, 55(3):1047-1063.

Houser, J.N., Mulholland, P.J. and Maloney. Catchment disturbance and stream metabolism: patterns in Ecosystem respiration and gross primary production along a gradient of upland soil and Vegetation disturbance. Journal of the North American Benthological Society. 24(3):538-552.

Horner, R.R., Welch, E.B. & Veenstra, R.B. 1983. Development of nuisance periphytic algae in laboratory streams in relation to enrichment and velocity. In: *Periphyton of Freshwater Ecosystems,* R.G. Wetzel (ed.), Dr. W. Junk Publishers, The Hague, The Netherlands: 121-131.

Howarth, R. W. & R. Marino. 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. *Limnology and Oceanography* 51(1, part 2): 2006, 364–376.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187-211.

Izagirre, O., Agirre, U. Bermejo, M., Pozoz, J. and Elosegi, A. 2008. Environmental controls of whole stream metabolism identified from continuous monitoring of Basque streams. *Journal of the North American Benthological Society*. 27(2):252.-268

Johnson, L.T., Tank, J.L. and Dodds, W.K. 2009. The influence of land use on stream biofilm nutrient Limitation across eight North American ecoregions. *Canadian Journal of Fisheries and AquaticSciences* 66:1081-1094.

Jones, J. B., J. D. Schade, S. G. Fisher, and N. B. Grimm (1997), Organic matter dynamics in Sycamore Creek, a Desert Stream in Arizona, USA, *Journal of North American Benthological Society*, *16*(1), 78-82.

Justus, B.G., Petersen, J.C., Femmer, S.R., Davis, J.V., and Wallace, J.E. 2010. A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams, Ecological Indicators 10(3): 627-638. http://dx.doi.org/10.1016/j.ecolind.2009.10.007

Kail, J., Arel, J. & Jahnig, S.C. 2012. Limiting factors and thresholds for macroinvertebrate assemblages in European rivers: Empirical evidence from three datasets on water quality, catchment urbanization and river restoration. *Ecological Indicators*. 18:63-72.

Kannel, P. R., S. Lee, Y. S. Lee, S. R. Kanel, and G. J. Pelletier (2007), Application of automated QUAL2Kw for water quality modeling and management in the Bagmati River, Nepal., *Ecological Modelling*, *202*(3-4), 503-517.

Kardouni, J., and N. Cristea (2006), Snoqualmie River Temperature Total Maximum Daily Load Study, edited by W. D. o. Ecology, Washington Department of Ecology, Olympia, WA.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries*. 6:21-27.

Kelly, O., K.O. Maloney, and D. E. Weller. 2011. Anthropogenic disturbance and streams: land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology* 56: 611–626.

Kemp, M.J. and Dodds, W.K. 2001. Centimeter-scale patterns in dissolved oxygen and nitrification rate in a prairie stream. *Journal of the North American Benthological Society*. 20(3):347-357.

(King et al. 2000)

King, R.S. & Richardson, C.J. 2003. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management*. 31(6):795-809.

King, R.S., Baker, M.E., Kazyak, P.F. and Weller, D.E. 2011. How novel is too novel? Stream community thresholds at exceptionally low levels of catchment urbanization. *Ecological Applications*. 21:1659-1678.

King, R.S. & Baker, M.E. 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. *Journal of the North American Benthological Society*. 29:998-1008.

King, R.S. & Richardson, C.J. 2004. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management*. 31:795-809.

Kiry, P., D. Velinsky, and Compton, A.M. 1999. Determination of dry weight and percent organic matter for sediments, tissues, and benthic algae. PCER Procedure P-16-113. Patrick Center forEnvironmental Research, Academy of Natural Sciences, Philadelphia.

Kominoski, J.S and Rosemond, A.D. 2012. Conservation from the bottom up: forecasting effects on global change on dynamics of organic matter and management needs for river networks. *Freshwater Science*. 31(1):51 68.

Lancaster, J. and B. J. Downes. 2010. Linking the hydraulic world of individual organisms to ecological processes: putting ecology into ecohydraulics. *River Research and Applications* 26:385-403.

Lane, C.S., Lyon, D.R. and Ziegler, S.E. 2012. Cycling of two carbon substrates of contrasting lability byheterotrophic biofilms across a nutrient gradient of headwater streams. *Aquatic Sciences*. Madsen, T.V. and Cedergreen, N. 2002. Sources of nutrients to rooted submerged macrophytes growing in a nutrient-rich stream. *Freshwater Biology*. 47:283-291.

Le, N. D. and J. V. Zidek. 1994. Network designs for monitoring multivariate random spatial fields. Pages 191-206 *in* J. P. Vilaplana and M. L. Puri, editors. Recent Advances in Statistics and Probability. VSP, Leiden, The Netherlands.

Lewis, Jr., W.M., Wurtsbaugh, W.A. and Paerl, H.W. 2011. Rationale for control of anthropogenic Nitrogen and phosphorus to reduce eutrophication of inland waters. *Environmental Science and Technology*. 45:10300-10305.

(Lewis et al. 1999).

Liess, A., Lange, K., Schulz, F., Piggott, J. J., Matthaei, C. D. and Townsend, C. R. (2009), Light, nutrients and grazing interact to determine diatom species richness via changes to productivity, nutrient state and grazer activity. Journal of Ecology, 97: 326–336.

Link, W. A., E. Cam, J. D. Nichols, and E. G. Cooch. 2002. Of bugs and birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management* 66:277-291.

Lovell, D. P. 2013. Biological importance and statistical significance. *Journal of Agricultural and Food Chemistry* 61:8340-8348.

(Lowe and Likens 2005).

Maine Department of Environmental Protection. 2009. Protocols for calculating the diatom total phosphorus index (DTPI) and diatom total nitrogen index (DTNI) for wadeable streams and rivers. DEPLW-0970A.

Maloney, K. O. and D. E. Weller. 2011. Anthropogenic disturbance and streams: land use and land use change affect stream ecosystems via multiple pathways. *Freshwater Biology* 56(3):611–626.

Marshall, B. D. 2009. Assessment of the biological condition of the New Fork River, in the vicinity of the Pinedale Anticline Project Area: 2008. Prepared for Sublette County Conservation District, Pinedale, WY.

Martinez-Abrain, A. 2008. Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology. *Acta Oecologica* 34:9-11.

Mason, S. J. Graham, G.E. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Quarterly Journal of The Royal Meteorological Society*. 128:2145-2166.

Mazumder, A. and Edmonson, J.A. 2002. Impact of fertilization and stocking on trophicinteractions and growth of juvenile sockeye salmon (Oncorhynchus nerku). Canadian Journal of Fisheries and Aquatic Sciences 59:1361-1373.

McBride, G. B., and S. Chapra (2005), Rapid calculation of oxygen in streams: approximate delta method, *Journal of Environmental Engineering*, 336-342.

McCune, B. 2011. Nonparametric multiplicative regression for habitat modeling. http://www.pcord.com/NPMRintro.pdf.

(McMormick et al. 2001)

McLaughlin, D.B. 2012. Assessing the predictive performance of risk-based water quality criteria using decision error estimates from receiver operating characteristics (ROC) analysis. *Integrated Environmental Assessment and Management*. 8(4):674-684.

McTammany, M.E., Benfield, E.F. and Webster, J.R. 2007. Recovery of stream ecosystem metabolism from historical agriculture. Journal of the North American Benthological Society. 26(3):532-545.

(Meritt et al. 2008)

Merow, C., M. J. Smith, T. C. Edwards, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wuest, N. E. Zimmermann, and J. Elith. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, doi: 10.1111/ecog.00845.

Meyer, S. M. 2004. End of the wild. “The extinction crisis is over. We lost”. Boston Review. April/May.

(Miltner and Rankin 1998).

Minshall, G. W. 1984. Aquatic insect-substratum relationships. Pages 358-400 *in* V. H. Resh and D. M. Rosenberg, editors. The Ecology of Aquatic Insects. Praeger Publishing, New York.

Minshall, G.W. 1978. Autotrophy in streams. *Bioscience*. 28(12):767-775.

Minnesota Pollution Control Agency. 2005. Minnesota lake water quality assessment report:developing nutrient criteria, Third Edition.

Minnesota Pollution Control Agency. 2008. Relation of nutrient concentrations and biological responses in Minnesota streams: applications for river nutrient criteria development.

Mitsui, A., S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao & T. Arai. 1986. Strategy by which nitrogen-fixing unicellular Cyanobacteria grow photoautotrophically. *Nature* 323: 720-722.

Montana Department of Environmental Quality. 2008. Scientific and technical basis of the numeric nutrient criteria for Montana.

Morgan, M. G. and M. Henrion. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, 6th edition. Cambridge University Press, 348 pp.

Mulholland, P.J., Houser, J.N. and Maloney, K.O. 2005. Stream diurnal dissolved oxygen profiles as indicators of in-stream metabolism and disturbance effects: Fort Benning as a case study. Ecological Indicators. 5:243-252.

Mulholland, P.J. Fellows, C.S., Tank, J.L., Grimm, N.B., Webster, J.R., Hamilton, S.K., Marti, E., Ashkenas, L., Bowden, W.B., Dodds, W.K., Mcdowell, W.H., Paul, M.J. and Peterson, B.J. 2001. Inter-biome Comparison of factors controlling stream metabolism. *Freshwater Biology* 46(11):1503-1517.

Mulholland, P.J., Fellows, C.S., Tank, J.L., Grimm, J.R., Webster, S.K., Hamilton, S.K., Marti, E., Ashkenas, L., Bowden, W.B., Dodds, W.K., McDowell, W.H., Paul, M.J. and Peterson, B.J. 2001. Inter biome comparison of factors controlling stream metabolism. *Freshwater Biology*. 46:433-451.

Mulholland, P.J., Stienman, A.D., Palumbo, A.V. and DeAngelis, D.L. 1991. *Journal of the North American Benthological Society* 10(2):127-142.

(Muotka and Laasonen 2002).

(Murray et al. 1987).

(Myer et al. 2007).

Naeem, S. 2002. Ecosystem consequences of biodiversity loss: the evolution of a paradigm. *Ecology*

Neilson, B.T., Hobson, A.J., von Stackelberg, N., Shupryt, M.P. and Ostermiller, J.D. 2012. Using Qual2k modeling to support nutrient criteria development and wasteload analysis in Utah. TechnicalReport to the Utah Department of Environmental Quality, Division of Water Quality, Salt LakeCity, Utah.

Nevers, M.B. & Whitman, R.L. 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. *Water Research.* 45(4):1659-1668.

Newbury, R. W. 1984. Hydrological determinants of aquatic insect habitats. Pages 323-357 *in* V. H. Resh and D. M. Rosenberg, editors. The Ecology of Aquatic Insects. Praeger Publishing, New York.

Nydick, K.R., B. Moraska, J. Lafrancois, S. Baron, and B.M. Johnson. 2004. Nitrogen regulation of algal biomass, productivity, and composition in shallow mountain lakes, Snowy Range, Wyoming, USA. *Canadian. Journal Fisheries and Aquatic Sciences* 61: 1256–1268.

Odum, H.T. 1956. Primary production in flowing waters. *Limnology and Oceanography.* 1:103-117.

Olsen. A.R. and Peck, D.V. 2008. Survey design and extent estimates for the wadeable streams assessment. *Journal of the North American Benthological Society.* 27(4):822-836.

(Olson and Hawkins 2013).

Opsahl, R.W., Wellnitz, T.W. and Poff, N.L. 2003. Current velocity and invertebrate grazing regulate Stream algae: results of an in situ electrical exclusion. *Hydrobiologia* 499:135-145.

Paasy, S.I. Continental diatom biodiversity in stream benthos declines as more nutrients become limiting. *Proceedings of the National Academy of Sciences.* 105:9663-9667.

(Paul and Myer, 2001)

Paulsen S.G., Mayio, A., Peck, D.V., Stoddard, J.L., Tarquinio, E., Holdsworth, S.M., Van Sickle, J., Yuan, L.L., Hawkins, C.P., Herlihy, A.T., Kaufman, P.R., Barbour, M.T., Larsen, D.P. and Olsen, A.R. 2008.Condition of stream ecosystems in the US: and overview of the first national assessment. *Journalof the North American Benthological Society.* 27(4):812-821.

Pelletier, G. J., and S. Chapra (2008), QUAL2Kw theory and documentation (version 5.1): A modeling framework for simulating river and stream water quality., edited by W. S. D. o. Ecology, Washington State Department of Ecology, Olympia, WA.

Pelletier, G. J., S. C. Chapra, and H. Tao (2006), QUAL2Kw - A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration., *Environmental Modelling & Software*, *21*(3), 419-425.

Peterson, B. J., L. Deegan, J. Helfrich, J. E. Hobbie, M. Hullar, B. Moller, T. E. Ford, A. Hershey, A. Hiltner, G. Kipphut, M. A. Lock, D. M. Fiebig, V. McKinley, M. C. Miller, J. R. Vestal, R. Ventullo, and G. Volk. 1993. Biological responses of a tundra river to fertilization. Ecology 74:653-672.

Pringle, C.M., Paaby-Hansen, P.D., Vaux, P.D. and Goldman, C.R. 1986. In situ assays of periphyton growth In a lowland Costa Rican stream. *Hydrobiologia.* 134:207-213.

Qian, S.S., King, R.S. and Richardson, C.J. 2003. Two statistical methods for the detection of environmental Thresholds. *Ecological Modeling* 166:87-97.

Quinn, G. and M. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press. 553 pp.

R Development Core Team. 2012. R: a language and environment for statistical computing.

Vienna,Austria. www/R-project.org Redfield, A.C. 1958. The biological control of chemical factors in the environment. *The American Scientist.* 46:205-221.

Regan, H. M., M. Colyvan, and M. A. Burgman. 2002*a*. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12(2):618–628.

Regan, H. M., Y. Ben-Haim, B. Langford, W. G. Wilson, P. Lundberg, and S. J. Andelman. 2005. Robust decision making under severe uncertainty for conservation management. *Ecological Applications* 15(4):1471-1477.

Redfield, A.C. 1958. The biological control of chemical factors in the environment. *The American Scientist.* 46:205-221.

Richards, D. C., T. Arrington, and W. VanWinkle. 2009. Metapopulation viability ananlysis and quantitative risk assessment of threatened snail in the Snake River, ID. Idaho Power Report to Federal Energy Regulatory Commission.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Muller, M. 2011. pROC: An open source package for R and S+ to analyze and compare ROC curves. *BMC Biometrics,* 12 p.77

Robinson, C.T. and Gessner, M.O. 2000. Nutrient addition accelerates leaf breakdown in an alpine spring brook. *Oceologia.* 122:258-263.

Rosemond, A.D., Mulholland, P.J. and Brawley. 2000. Seasonally shifting limitation of stream periphyton: Response of algal populations and assemblage biomass and productivity to variation in light, nutrients and herbivores. *Canadian Journal of Fisheries and Aquatic Sciences 57:66-75*

Sabater, S., Gregory, S.V. and Sedell, J.R. 2002. Community dynamics and metabolism of benthic algae Colonizing wood and rock substrata in a forested stream. *Journal of Phycology*. 34(4):561-567.

Sanderson, B.L., Coe, H.J., Tran, C.D., Macneale, K.H., Harstad, D.L. and Goodwin, A.B. 2009. Nutrient Limitation of periphyton in Idaho streams: results from nutrient diffusing substrate experiments. *Journal of the North American Benthological Society* 28(4):832-845.

Schindler, D. W., F. A. J. Armstrong, S. K. Holmgren, and G. J. Brunskill. 1971. Eutrophication of Lake 227, Experimental Lakes Area, Northwestern Ontario, by Addition of Phosphate and Nitrate. *Journal of the Fisheries Research Board of Canada* 28(11):1763-1782.

Schindler, D.W. 1977. Evolution of phosphorus limitation in lakes. *Science* 195:260-262.

(Scott et al. 2008)

Scott, J.H., Lang, D.A., King, R.S. and Doyle, R.D. 2009. Nitrogen fixation and phosphatase activity in periphyton growing on nutrient diffusing substrata: evidence for differential nutrient limitation In stream periphyton. *Journal of the North American Benthological Society*. 28(1):57-68.

Seife, C. 2010. Proofiness: the Dark Arts of Mathematical Deception. Penguin Books, New York, NY.

Slavik, K., B. J. Peterson, L. A. Deegan, W. B. Bowden, A. E. Hershey, and J. E. Hobbie. 2004. Long-term responses of the Kuparuk river ecosystem to phosphorus enrichment. Ecology 85:939-95.

Smith, V. H. and D. W. Schindler. 2009. Eutrophication science: where do we go from here? *Trends in Ecology and Evolution* 24:201-207.

Smith, A.J., Bode, R.W. & Kleppel, G.S. A nutrient biotic index (NBI) for use with benthicmacroinvertebrate communities. *Ecological Indicators*. *7*(2):371-386.

(Smith et al. 2003).

(Sumberkropp et al. 2010).

Suplee, M.W., Watson, V., Teply, M. & McKee, H. 2009. How green is too green? Public opinion of what constitutes undesirable algae levels in streams. *Journal of the American Water Resources Association*. *45*(1): 134-140.

(Suplee et al. 2012)

(Suplee and Watson 2013).

Stantec Consulting (2010), Jordan River TMDL: 2010 QUAL2Kw Model Calibration Technical Memo PUBLIC DRAFT. , edited by D. o. W. Q. Utah Department of Environmental Quality, Salt Lake City, UT.

Statzner, B. and L. A. Beche. 2010. Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology*. *55*: 80-119.

(Stelzer and Lamberti 2001)

Stelzer, R.S., Heffernan, J. and Likens, G.E. 2003. The influence of dissolved nutrients and particulate organic matter quality on microbial respiration and biomass in a forest stream. 48:1925-1937.

(Stoner 2011).

Tank, J.L., Rosi-Marshall, E.J., Griffiths, N.A., Entrekin, S.A. and Stephen, M.L. 2010. A review of allochthonous organic matter dynamics and metabolism in streams. *Journal of the North American Benthological Society*. 29(1):118-146.

Tank, J.L. and Dodds, W.K. 2003. Nutrient Limitation of epilithic and epixylic biofilms in ten North American streams. *Freshwater Biology* 48:1031-1049.

Tank, J.L., Bernot, M.J. and Rosi-Marshall, E.J. 2006. Nitrogen limitation and uptake. In Hauer, F.R. Lamberti, G.A. Editors, *Methods in Stream Ecology*, 213-238. Elsevier, New York.
Valderamma, J.C. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. *Marine Chemistry*. 10(2):109-122.

Tank, J.L. and Webster, J.R. 1998. Interaction of substrate and nutrient availability on wood biofilm processes in streams. *Ecology*. 79:2168-2179.

Thompson, S. K. and A. F. Seber. 1996. Adaptive Sampling. Wiley, New York, NY, USA.

Tockner K., M. Pusch, D. Borchardt and M.S. Lorang. 2010. Multiple stressors in coupled river–floodplain ecosystems. *Freshwater Biology* *55*(Suppl. 1): 131–151.

Townsend, C.R., S.S. Uhlmann and C.D. Matthaei. 2008. Individual and combined responses of stream ecosystems to multiple stressors. *Journal of Applied Ecology* *45*: 1810–1819.

Townsend C.R.and A.G. Hildrew. 1994. Species traits in relation to a habitat template for river systems. *Freshwater Biology* 31: 265–275.

Triska, F.J. and Sedell, J.R. 1976. Decomposition of four species of leaf litter in response to nitrate manipulation. *Ecology*. *57*:783-792.

Uehlinger, U. Kawecka, B. and Robinson, C.T. 2003. Effects of experimental floods on periphyton and stream metabolism below a high dam in the Swiss Alps (River Spol). *Aquatic Sciences*. 65(3):199-209

UNEP 1992. Rio Declaration on Environment and Development. (http://www.unep.org/Documents.multilingual/Default.asp?DocumentID=78&ArticleID=1163)

U S Environmental Protection Agency. 2012. Water Quality Standards Handbook: Second Edition. EPA-823-B-12-002.
U.S. Environmental Protection Agency, Washington DC.

U S Environmental Protection Agency. 2010. Using stressor-response relationships to derive numeric nutrient criteria*Rep.*, U. S. Environmental Protection Agency, Washington, D.C.

U S Environmental Protection Agency. National Rivers and Streams Assessment: Field Operations Manual. EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

U S Environmental Protection Agency. National Rivers and Streams Assessment: field operations manual. EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

U S Environmental Protection Agency. Wadeable streams assessment: A collaborative survey of the Nation's streams. EPA 841-B-06-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

U S Environmental Protection Agency. Summary of Biologic Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington D.C.

U S Environmental Protection Agency. 2000. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

U S Environmental Protection Agency. 1986. Quality Criteria for Water*,* Office of Water, EPA 440/5-86-001, Washington, D.C.

U S Environmental Protection Agency. 1976. Quality criteria for water. EPA-440/9-76-023. Washington, DC.

Utah Division of Water Quality. 2011. Field data collection for QUAL2Kw model build and calibration. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, Utah.

Vallentyne, J.R. 1974. The algal bowl—lakes and man. Ottawa, Misc. Special Publ.22, Dept. of the Environment, 185pp.

Valderrama, J.C. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. *Marine Chemistry*. 10(2):109-122.

Van de Bogert, M.C., Carpenter, S.R., Cole. J.J. and Pace, M.L. 2007. Assessing pelagic and benthic metabolism using open channel measurements. *Limnology and Oceanography:Methods*. 5:145-155.

Vanote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. and Cushing, C.E. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*. 37:130-137.

(Vanote et al. 1980).

Van Sickle, J. and Paulsen, S.G. 2008. Assessing the attributable risks, relative risks and regional extents of aquatic stressors. *Journal of the North American Benthological Society*. 27(4):920 931

Van Sickle, J., Stoddard, J.L., Paulsen, S.G. & Olsen, A.R. 2006. Using relative risk to compare the effects of aquatic stressors at a regional scale. *Environmental Management*. 38:1020-1030.

Vieira, N. K. M., N. L. Poff, D. M. Carlisle, S. R. Moulton II, M. K. Koski, and B. C. Kondratieff. 2006. A database of lotic invertebrate traits for North America: U.S. Geological Survey Data Series 187. Available at: http://pubs.water.usgs.gov/ds187.

Vollenweider, R. D. 1968. The scientific basis for eutrophication of lakes and streams, with particular reference to phosphorus and nitrogen as eutrophication factors. Technical Report OECD. Paris. DAS/CSI/68, 27:1182.

(Von Shiller et al 2000)

Wagenhoff, A., C. R. Townsend, and C. D. Matthaei. 2012. Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. *Journal of Applied Ecology* 49(4):892–902.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss. 2003. Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* 4(1):5–17.

Walther, D.A. and Whiles, M.R. 2011. Secondary production in a southern Illinois headwater stream: relationships between organic matter standing stocks and macroinvertebrate productivity. 30(2):357-373.

(Wallace and Webster 1996)

(Wallace et al. 1982)

Water Studies Centre Monash University, Murray Darling Basin Commission and New SouthWales Department of Environment and Climate Change. 204pp

Wang, L., Robertson, D.M. & Garrison, P.J. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. *Environmental Management*. 39:194-212.

Webster, J.R. and Meyer, J.L. 1997. Organic matter budgets for streams: a synthesis. *Journal of the North American Benthological Society*. 16:141-161.

Webster, J.R., Golladay, S.W., Benefield, E.F., D'Amgelo, D.J., Peters, G.T. 1990. Effects of forest disturbance on particulate organic matter budgets of small streams. *Journal of the North American Benthological Society*. 9:120-140.

Webster, J. R., J. B. Waide, and B. C. Patten. 1975. Nutrient cycling and the stability of ecosystems. Pages 1-27 *in* F. G. Howell, J. B. Gentry, and M. H. Smith, editors. Mineral Cycling in

Southeastern Ecosystems. ERDA CONF-740513. Reprinted as pages 136-162 *in* Systems Ecology. H. H. Shugart and R. V. O'Neill, editors. Benchmark Papers in Ecology. Dowden, Hutchinson, and Ross Inc., Stroudsburg, PA.

Weigel, B.M. & Robertson, D.M. 2007. Identifying biotic integrity and water chemistry relations in nonwadeable rivers of Wisconsin: toward the development of nutrient criteria. *Environmental Management*. 40(4):691-708.

Wetzel, Robert G. *Limnology*. 1975. pp. 520-523.

Wikle, C. K. and J. A. Royle. 1999. Space-time models and dynamic design of environmental monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics* 4:489-507.

Wikle, C. K. and J. A. Royle. 2005. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics* 16:507-522.

Yamamuro, A.M and Lamberti, G.A. 2007. Influence of organic matter on invertebrate colonization of sand substrata in a northern Michigan stream. *Journal of the North American Benthological Society*. 26(2):244-252.

Ylla, I., Romani, A.M. and Sabater, S. 2012. Labile and recalcitrant organic matter utilization by river biofilm under increasing water temperature. *Microbial Ecology*. 64(3):593-604.

Young, R.G., and Huryn, A.D. 1999. Effects of land use on stream metabolism and organic matter turnover. *Ecological Applications*. 9:1359-1376.

Young, R., Townsend, C. & Matthaei, C., 2004, Functional indicators of river ecosystem health. Cawthorn Institute. http://www.cawthron.org.nz/media_new/publications/pdf/2014_07/FunctionalIndicatorsGuideFinal_2.pdf

Young, R.G., Matthaei, C.D. and Townsend, C.R. 2008. Organic matter breakdown and ecosystem metablism: functional indicators for assessing river ecosystem health. *Journal of the North American Benthological Society*. 27(3):605-625.

Ziliak, S. T. and D. N. McCloskey. 2009. The cult of statistical significance. Pages 2302-2316 *in* Proceedings of the Joint Statistical Meetings, Section on Statistical Education, 3Aug2009, Washington, DC.

Zimmerman, D. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* 17:635-662.

Appendices

APPENDIX A: STANDARD OPERATING PROCEUDRES

Nutrient Diffusing Substrates

Placeholder

We have drafts of these and can provide them on request, but we're working on converting them to a format consistent with DEQ QAPPs for the final report.

Whole Stream Metabolism

Placeholder

We have drafts of these and can provide them on request, but we're working on converting them to a format consistent with DEQ QAPPs for the final report.

Organic Matter Standing Stocks

Placeholder

We have drafts of these and can provide them on request, but we're working on converting them to a format consistent with DEQ QAPPs for the final report.

Synoptic Sampling Procedures for Purposes of Model Calibrations

Placeholder

We have drafts of these and can provide them on request, but we're working on converting them to a format consistent with DEQ QAPPs for the final report.

APPENDIX B: ADDITIONAL MATERIAL FOR CHAPTER 11

Placeholder

In review.

APPENDIX C: ADDITIONAL MATERIALS IN SUPPORT OF MECHANISTIC MODELING

This section contains several reports that were submitted in conjunction with the modeling component of the nutrient study. The studies primarily speak to the integration of process-based models in the creation of site-specific criteria, but some information is of broad interest.

A DATA COLLECTION AND CALIBRATION STRATEGY FOR QUAL2Kw

A.J. Hobson[1], B.T. Neilson[2], N. von Stackelberg[3], M. Shupryt[4], J. Ostermiller[5], G. Pelletier[6], S. C. Chapra[7]

1. Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University

2. Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University

3. Utah Department of Environmental Quality, Division of Water Quality

4. Utah Department of Environmental Quality, Division of Water Quality

5. Utah Department of Environmental Quality, Division of Water Quality

6. Washington State Department of Ecology

7. Civil and Environmental Engineering, Tufts University

Abstract

In-stream water-quality models provide guidance in watershed management decisions by linking pollutant loads to changes in water quality. These models are particularly useful for determining wasteload allocations, developing numeric nutrient criteria, and aiding in total maximum daily load (TMDL) analyses. Unfortunately, the routine data collected as part of the governmental monitoring efforts do not typically meet the data requirements for modeling. Consequently, this study presents a foundational data collection methodology suited to meet in-stream water-quality modeling requirements for a commonly used model (QUAL2Kw). To set some model parameters directly, methods are provided for estimating maximum sediment oxygen demand and appropriate reaeration formulas using observed oxygen time series. The quantity of many data types was minimized to reduce cost which resulted in challenges due to data limitations (e.g., designation of appropriate loading values from highly variable point source information). Similar to other modeling studies, parameter estimates were also not readily identifiable. However, even simple methods to reduce the number of parameters requiring calibration proved beneficial. Although most problems will require additional model calibration and data for model corroboration, this approach provides an initial framework that aids in the judicious use of resources to meet watershed management decision making needs within the context of wasteload allocation and/or numeric nutrient criteria development.

Introduction

In-stream water-quality models can be helpful in the watershed management decision process by understanding nutrient loading effects on changes in water quality (Boyacioglu and Alpaslan 2008; Kannel et al. 2011; National Research Council 2007; Orlob, 1992; von Stackelberg and Neilson 2014). Such models are used for a variety of applications including wasteload allocations (WLAs) (UDEQ 2012b), establishing regional or site-specific numeric nutrient criteria (NNC) (Flynn and Suplee 2011; US EPA 2000), and total maximum daily load (TMDL) assessments (Boyacioglu and Alpaslan 2008; National Research Council 2001). Many of these applications focus on critical low-flow periods (Bischoff et al. 2010; Gunderson and Klang 2004; Stahl and Smith 2002; UDEQ 2000; US EPA 2002a) that result in high primary productivity, low dissolved oxygen (DO) levels, and elevated stream temperatures (US EPA 1997). These conditions often exceed in-stream water-quality standards, approaching thresholds of many aquatic organisms (Hester and Doyle 2011), and are only expected to worsen in the future with global climate change (Whitehead et al. 2009). During critical periods in riverine systems, simplified one-dimensional, quasi-dynamic (constant flow with diel weather and water quality) models, such as QUAL2E (Brown and Barnwell 1987) and QUAL2K (Chapra et al. 2006), are typically employed to represent the fate and transport of solutes in the downstream direction (US EPA 1997). A modified version of these models maintained and distributed by Washington State Department of Ecology, QUAL2Kw (Pelletier et al. 2006), has been selected for water-quality impairment assessments conducted by many state and national agencies (Carroll et al. 2006; Kannel et al. 2011; Turner et al. 2009).

All models require data input for model setup, including physical characteristics (hydraulic information and channel segmentation), forcing (meteorological, boundary conditions, and point and distributed sources) and calibration data (in-stream observations) to adequately characterize effects on water quality from significant loading sources. The supporting data collection campaigns must capture stream variability (both temporal and spatial) and necessary data types while often adhering to stringent budget requirements (Neilson and Chapra 2003; US EPA 2002b). As the need for WLAs, TMDLs, and site-specific NNC causes a burden on internal resources for public and private agencies (Lettenmaier et al. 1991), a reliable and systematic method of collecting data to support in-stream modeling is needed. Sampling strategies must be established that limit the number of measurements and data types collected without having a consequential impact on model reliability (Dunnette 1980; Facchi et al. 2007; Henderson-Sellers and Henderson-Sellers 1996).

To address the need for protocols to optimize the allocation of limited resources, this paper presents a systematic and foundational data collection, model setup, and model calibration framework for applying QUAL2Kw to riverine systems impacted by water reclamation facilities (WRF). Assuming steady state flow conditions, the generalized approach provides guidance for basic data collection given the temporal and spatial variability of point source impacted reaches, as well as options for additional data collection to support model parameterization and calibration. To evaluate the effectiveness and limitations of the generalized and basic approach, a case study is presented of a model application to an effluent-dominated headwater stream in Utah.

Generalized Data Collection and Modeling Approach

The generalized approach was initially developed and applied to 6 study sites throughout central and northern Utah (SI Figure 1). Throughout this process, site specific hydrologic characteristics, sampling problems (e.g., missing samples), and varying influences of the WRFs provided insight regarding data requirements and provided the basis for the sampling plan presented here. This generalized data collection and modeling framework relies upon synoptic surveys of point source impacted reaches to support model setup and calibration. Synoptic surveys require concurrent sampling of the effluent as well as upstream and downstream locations. These surveys are especially important during the critical period, which is typically summer low-flow conditions for this type of system (e.g., Turner et al., 2009). However, additional synoptic surveys collected under various environmental conditions would serve for model confirmation after calibration, and also to define the temporal extent of the critical season to guide seasonal wastewater treatment requirements.

Sampling Locations **-** To model a study reach (Figure 1), data must be collected to capture the variability of the headwater (also called the upstream boundary condition, Station 1), point sources or abstractions (e.g., WRF, tributary inflows, irrigation diversions at Station 2, T1 and D1, respectively) and any diffuse sources or abstractions (e.g., groundwater). The type of information that should be gathered at each station varies (Figure 1, Table 1). In the context of effluent dominated systems, at an absolute minimum, supporting data need to be gathered at the headwater/upstream boundary condition (Station 1), point source before it enters the stream (Station 2), and the downstream end of the study reach for calibration (Station 3). However, placing additional stations at the initial mixing point of a WRF and at or beyond the point of maximum impact of the point source is desirable. If a significant tributary (e.g., greater than 10% of the study reach flow (Bartholow 1989)) enters the modeling reach, flow and quality data must also be collected at Station T1. Also, if a significant diversion is present, flow information is needed at Station D1 (the water quality of the diversion is the same as that in the modeled reach).

Once the stations are identified, the distance between Stations 2 and 3 can be determined using various methods and informed by site-specific criteria, but should located downstream of the mixing zone and be long enough to capture the processes of interest. In general, selection of the Station 3 location must balance the need to capture the maximum effect of the discharge while minimizing confounding factors of tributaries, diversions, and groundwater. Further, it may be appropriate to select the distances based on requirements to derive estimates of open-water metabolism and surface reaeration (k_a) which may require additional intermediate stations.

Data Types **-** Data are required for a number of water-quality constituents at each station. The requirements are dependent on whether it is the headwater station, a point or distributed inflow, or a diversion (Figure 1, Table 1). Data collected at each station should represent the study period of interest (e.g., summer low flow conditions). Some constituents can be sampled directly while others are estimated using relationships between measured constituents and model variables (SI Table 1). An estimate of bottom or benthic algae concentrations should also be measured, particularly in shallow streams or rivers. Additional data types that could be collected include a measure of sediment oxygen demand (SOD), total organic carbon, and dissolved organic carbon (DOC) (used to estimate CBOD and/or detritus).

Beyond water-quality data, site-specific information is necessary to characterize the stream and its surroundings. This includes geometric (bottom slope, channel cross-sections), meteorological (air temperature, dew point temperature, wind speed, and cloud cover or solar radiation), and hydraulic (travel time, stream and groundwater flow, velocity, substrate types, and percent suitable substrate) information. A summary is provided with a list of the data types to collect, some procedural information, locations where these data are required within or near the site, and the utility of the data in the context of the modeling effort (Table 2).

Sampling Frequency **-** When one considers the timing of sample collection in the context of point source impacted reaches, it can be difficult to resolve whether observed diel fluctuations are from point source variability or simply an artifact of sampling time and background diel fluctuation cycles (Nimick et al. 2011). When possible, site-specific information on spatial and temporal variations of specific water-quality constituents should be obtained prior to committing to a sampling plan (Ort et al. 2010). For example, pre-sampling reconnaissance can include deployment of continuous sondes to explore the variability and timing of diel minimum and maximum conditions at various locations. In general, those constituents that have the highest variability require the most frequent sampling interval. Some constituents that can be measured *in situ* using multi-parameter sondes (e.g., temperature, DO, conductivity) can be sampled most frequently during the study period. Although these sensors are typically limited more by cost than by temporal sampling frequencies, measurements made hourly for at least a 24-hour period, and preferably over 2-3 days or the duration of the synoptic sampling event, should adequately capture a typical diel signal and provide appropriate estimation of constituents for modeling needs (Gammons et al. 2011). Grab sample frequency of the remaining constituents requiring laboratory analyses are typically limited by personnel and cost. Therefore, entities commonly rely on intermittent sampling due to assumed low diel variability or on historic values for modeling applications (Bischoff et al. 2010; Carroll et al. 2006). Sampling frequency has been studied extensively (Facchi et al. 2007; Fogle et al. 2003; Hazelton 1998; Henjum et al. 2010; Ort et al. 2010; Zhang and Zhang 2012) and some guidance on choosing a sampling strategy is offered specifically for applications in general TMDL analyses and WLAs (US EPA 1986, 1995, 2002b). If possible, sampling should occur at least twice a day to target the times of expected diel minima and maxima of various constituents (e.g., at dawn and after solar noon or dusk) during the beginning and end of the study period at all key stations (Chapra 2003).

Model Setup - Once these water-quality and supplemental data are collected, model setup requires translation of stream observations to the input format requirements of the model. First, the study reach must be segmented and information regarding the channel geometry of each model reach must be determined from the observations. Next, all flow records from each station should be averaged to provide a single flow value for each point source/diffuse inflow, tributary, diversion, and the headwater. Then, measured water-quality data needs to be averaged to produce hourly estimates (headwater), or summary statistics (average, min, max, time of max concentration for assumed sine curve) for point source loads. In the case of limited data availability, as is common among chemistry and nutrient sampling, values can be averaged to provide daily mean concentrations and applied as a single value that does not have diel variation.

Another consideration with limited data availability occurs when chemistry and nutrient samples are analyzed and reported at very low concentrations that result in censored values or samples reported as below analytical or method detection limits (MDL). These cases require consideration of appropriate methods for estimating the true values, since either omitting a censored value, replacing it with zero, 0.5 MDL, or MDL will affect estimates of the mean, median, and variance of the observations. In cases where censored data constitute greater than 25% of the sample size, the selection of an appropriate method becomes more arbitrary (Berthouex and Brown 2002).

Model Calibration - After data collection and model setup, parameter values need to be set to accurately predict site-specific responses. Parameters are often established on the basis of the modeler's experience from applications in other systems, trial and error, or with optimization algorithms (Scholten and Refsgaard 2010). This is important because parameters that are set based on measurements (direct or indirect (Barnwell, Brown et al. 2004)) are typically more accurate than those estimated through calibration (Hattermann et al. 2010). While it is recognized that a calibration approach for parameter selection can be problematic (Guadagnini and Neuman 1999), improving data shortfalls can reduce cases where multiple parameter combinations produce the same water-quality predictions (equifinality) (Ebel and Loague 2006).

Model calibration should begin by establishing that certain constituents are predicted correctly before moving onto the more interconnected mechanisms associated with nutrient cycling. First, the flow balance and hydraulics should be verified so that the representation of the residence time and volumes are appropriate. Predicted discharge at downstream locations must match observations. If the values differ substantially, it could be due to inflows or outflows from unknown sources or from groundwater exchanges. These types of sources can be identified using simple differential gauging methods (Ruehl et al. 2006) to provide net changes in flow at various locations throughout a reach. The resulting gains and losses can be assumed to be a distributed groundwater source or abstraction. Abrupt changes in the longitudinal profile of specific conductance from upstream conditions can also provide supporting evidence of the presence of groundwater inflow (Cirpka et al. 2007; Vogt et al. 2010). Once the locations of inflows are roughly identified, the corresponding constituent concentrations require estimation or measurement and can be obtained from nearby seeps (groundwater that surfaces prior to the stream) or shallow groundwater observation wells (Covino and McGlynn 2007; Harvey et al. 1996).

Hydraulic geometry may be specified using either exponential rating curves of velocity (*U*) and depth (*Y*) versus flow (*Q*) (e.g., $Y = a\,Q^{b}$ and $U = c\,Q^{d}$), or Manning's equation. The coefficients (a and c) and exponents (c and d) for rating curves can be estimated from either long-term gauging station records or hydraulic models (e.g., HEC-RAS). To ensure appropriate flow routing, travel times must be validated. Travel times within the study reach are dependent on hydraulic geometry. If Manning's is used, then Manning's n is typically adjusted to calibrate the depth and velocity since width, channel slope, and side slope should be measured or a rectangular channel assumed. Tracer studies can be helpful in providing data to estimate travel times within the study reach, which in turn, can help gauge the accuracy of average estimates of bottom width, bottom slope, and side slope values.

Longitudinal and diel temperature predictions at different sub-reaches can primarily be adjusted through topographic and riparian shading estimates. Necessary shading information can be estimated using various methods [i.e., SHADE model (Chen 1996)] at each reach element by designating the nearest topographical feature (north, east, and west coordinates and % inclination), vegetation type, and the distance from stream center to the edge of the riparian zone. These types of tools can be used to estimate the hourly percent shading values required by QUAL2Kw. Another consideration is the accuracy of the predicted hydraulic geometry because water temperature response to variations in surface heat fluxes are very sensitive to geometry. At times, it may be necessary to revisit the channel geometry estimates to ensure the accuracy of temperature predictions.

Next, inorganic suspended solids (ISS) settling rate regulates the amount of suspended sediments in the water column, which is important for simulating light penetration through the water column. It can be set directly by adjusting the settling rate to calibrate the ISS predictions to observed stream conditions.

Finally, k_a and SOD can be estimated prior to calibration using various "open-water" methods of determining ecosystem metabolism. A general approach to using ecosystem metabolism methods follows that a change in oxygen over time (dO/dt) is a result of oxygen sources (primary production and reaeration) and oxygen sinks (autotrophic and heterotrophic respiration, BOD, and other oxygen consuming reactions within the water column and sediments); however, the relationships describing the change in oxygen is often reduced to Eq. 1:

$$\frac{dO}{dt} = k_a D + GPP - ER \qquad \textbf{(8)}$$

where k_a = stream reaeration rate (d^{-1}), D = DO deficit ($O_{sat} - O$) (mg L^{-1}), GPP = gross primary production (mg O_2 L^{-1} d^{-1}), and ER = ecosystem respiration (mg O_2 L^{-1} d^{-1}).

Some examples of using this relationship in a stream metabolism context have been established with the Delta Method (Chapra and Di Toro 1991; McBride and Chapra 2005), Nighttime Regression Method (Young et al. 2004), and Inverse Method (Holtgrieve et al. 2010) which simultaneously estimate k_a, GPP, and ER from the diurnal signal of DO from a single stream station (Eq. 1). With these values established at different points within a study reach, assuming k_a approximations are reasonable, these values can be used to determine which of the widely recognized reaeration formulas provided within QUAL2Kw may be appropriate to predict and represent reaeration under different flow conditions. If diel data are collected from several different river flow conditions then it is possible to derive a site-specific equation to estimate k_a from velocity and depth (e.g., solve for a, b, and c in an equation of the form $k_a = aU^{b}Y^{c}$). Otherwise, we suggest running the model using each reaeration formula and comparing predictions against the point estimates of k_a from stream metabolism methods. The most appropriate formula within QUAL2Kw can then be selected based on a combination of criteria (e.g., lowest root mean square error, RMSE), appropriateness of formula restrictions or assumptions).

QUAL2Kw has the functionality to estimate SOD based on a sediment diagenesis algorithm (Di Toro, Paquin et al. 1991; Di Toro and Fitzpatrick 1993; Di Toro 2001), but there is often more SOD present than is predicted due to the deposition of organic matter before the time period of model simulation (i.e., during snowmelt runoff) and the deposition of coarse particulate organic matter (CPOM) that typically is not captured by standard sampling techniques. Beyond sediment diagenesis, an additional amount of SOD can be prescribed within the modeling framework but it is handled as a direct sink of oxygen. Since site-specific or reach-integrated SOD measurements are generally not available, there is a need to approximate a reasonable reach scale SOD for each study site. An approach to estimating a maximum SOD is by subtracting *GPP* from *ER*, as defined in Eq. 1. This requires the assumption of autotrophic respiration approximately equaling *GPP* [it may need to be some fraction of *GPP* (Jones et al. 1997)] and any extra oxygen consumption is due to heterotrophic respiration and other oxygen consuming reactions within the sediments and water column. Based on this assumption, a positive value (meaning *ER* is higher than primary production) provides an estimate of total SOD (heterotrophic respiration + oxygen demanding reactions within the sediments) and some oxygen demanding reactions within the water column (e.g., BOD and nitrification). Within QUAL2Kw, it can be assumed that this total SOD would provide a maximum value that includes the prescribed SOD plus the SOD estimated within the sediment diagenesis algorithm [described within (Pelletier and Chapra 2008)]. The assumption that *ER* minus *GPP* equals SOD is assumed appropriate in typical effluent-dominated streams that are relatively shallow and where sediment processes significantly influence the water column DO response. In some situations it is possible that other processes have a more dominant influence on the water column oxygen responses (e.g., chemical reactions within the water column) and these approaches may not be applicable or include more error due to the aforementioned assumptions. Where appropriate, these SOD estimates can provide an upper bound to be used in calibration or an average reach value of SOD could be established and set before auto-calibration.

With a number of parameters set either from direct/indirect measurements or based on these prior manual calibration steps, the remaining parameters that are appropriate to include in model calibration can be auto-calibrated. Using the PIKAIA genetic algorithm (Charbonneau and Knapp 1995) within QUAL2Kw, the number of model runs over which to perform the optimization of the parameter set can be selected (# Model Runs = # of Populations × # of Generations). The parameters that are commonly included in auto-calibration as well as some appropriate parameter ranges are identified in SI Table 2. Within the auto-calibration, a fitness statistic is evaluated for desired state variables as the reciprocal of a weighted average of the normalized RMSE (Pelletier et al. 2006). This tool allows the coefficient of variation of the RMSE between each constituent (model results versus observed data) along with individual weighting factors, to be summarized in a single value that the genetic algorithm seeks to maximize by adjusting all desired parameters.

Case Study

The data collection, model setup, and calibration strategies presented above provide a general framework that can be adapted for different applications of QUAL2Kw or similar models. Once developed and refined, the finalized data collection and modeling approach was applied to a reach in Silver Creek, Utah during low-flow conditions in order to test the validity of the minimal data collection proposed. Assuming constant flow conditions, the data collection prescribed in the generalized approach was used to set up and calibrate the model, with additional data (e.g., SOD measurements and bottom algae samples) collected to evaluate model performance under such conditions.

Silver Creek is a small tributary to the Weber River with land use comprised mainly of Park City, two ski resorts and grazing. The study reach is located 6 miles north of Park City and is approximately 2 km in length near the middle of a 103 km^2 watershed (Figure 2). The climate is typical for high elevation, western mountainous regions, with the majority of the annual precipitation load attributed to winter snowfall and subsequent spring runoff (Whitehead and Judd 2004). During the summer months, some or all of the flow in Silver Creek is diverted upstream of the study reach for irrigation and stock watering purposes, therefore, Silver Creek becomes highly effluent-dependent downstream of the WRF.

Data Collection - The Silver Creek WRF is the major point source for nutrient loading to the Silver Creek study reach although various surface and groundwater seeps also contribute loads (Figure 2 shown as S1, S2, and S3). A small tributary, T1 (Figure 2b), also enters the stream reach one km downstream of the headwater (Station 1). The distance between Stations 2 and 3 was determined according to guidelines set by Grace and Imberger (2006) which designate station spacing based on reaeration estimates derived from depth and velocity measurements of the stream. To verify in-stream processes, intermediate measurement stations were established to provide more detailed hydraulic and water-quality data and are labeled USGS, I1, and I2 at river kilometers 1.8, 1.1 and 0.8, respectively. In addition, a time of travel study between Stations 2 and 3 was conducted using salt (NaCl) as a tracer. Travel time had to be estimated from mean velocity values between Stations 1 and 2 due to low channel flows and large pools, causing a tracer response to become indeterminable at Station 2.

In situ multi-parameter data loggers (YSI 6690 V2, Yellow Springs Instrument Company, Yellow Springs, OH) were deployed from August 22-30, 2011 at Stations 1, 2, T1, and 3 (Figure 2) to collect continuous diel data (five minute intervals) for DO, DO saturation, temperature, pH, conductivity, and chlorophyll *a*. Intermediate stations USGS, I1, and I2 logged from 5 to 15 minute intervals for DO, temperature, conductivity, and pH. YSI protocols were followed for sensor calibration that included a pre-deployment check with all sensors logging in ambient water for 30 minutes prior to deployment and a post-deployment check conducted in the same manner.

Chemistry and nutrient grab samples were collected at Stations 1, 2, T1, and 3 taken twice a day over a two day sampling period per the minimum data requirements protocol. Sampling times for the nutrient and chemistry data were chosen, without prior detailed sampling, to represent the assumed diel variation of constituents governed by the photoperiod with a dawn sample the first day and an afternoon sample the second day (Chapra 2003). Surface seeps S1, S2, and S3 were sampled once during the entire period.

The grab samples were collected according to operating procedures developed by the Utah Division of Water Quality (UDEQ 2012a). They were analyzed for $sCBOD_5$, total nitrogen, total dissolved nitrogen, ammonium, nitrate + nitrite, total phosphorus, total dissolved phosphorus, soluble-reactive phosphorus, chlorophyll *a*, pH, alkalinity, total suspended solids, and volatile suspended solids according to standard methods. From these measured constituents, others were calculated including organic nitrogen, organic phosphorus, detritus and inorganic suspended solids (SI Table 1). Additional samples were collected from the effluent of the WRF to provide an estimate of CBOD oxidation rates. These rates were estimated following EPA method 405.1 (including a nitrification inhibitor) by measuring DO in six reactors each day for 30 days to obtain an average CBOD oxidation rate of 0.103 d^{-1}. Finally, bottom algae estimates from a study conducted in September 2011 produced areal estimates at Stations 1, I2, and 3 and were used to check calibration performance. Values provided are a total of the scaled dry mass that considers the fractional cover of various channel and habitat types for each study section. Macrophyte coverage was omitted from the reported values.

In addition to water-quality samples, supplementary data were collected consisting of geometric, hydraulic, meteorological, and shading information. Width, depth, velocity, and flow measurements were taken several times at each station (S1, S2, T1, and S3) and intermediate stations (I1, I2) along with the high frequency flow record available from the USGS station (USGS 10129900, Silver Creek near Silver Creek Junction, UT). Due to large uncertainties in the flow data, a flow balance study was used to quantify the sources and sinks beyond the major observable inflows in the reach. High variability in channel geometry at Station 1 resulted in inaccurate estimates and required flow to be back-calculated from high frequency flow records from the USGS station and WRF ($Q_{HW} = Q_{USGS} - Q_{WRF}$). Additional distributed inflows and abstractions were quantified based on a channel water balance conducted using differential gauging methods and accounting for known inflows. Meteorological information was downloaded from two weather stations within 25 miles of Silver Creek (National Weather Service stations UTSVC and UTQRY). Data from these stations were used to provide hourly air temperature, dew point temperature, solar radiation, and wind speed during the study period. Longitudinal shading was derived from the SHADE model (Chen 1996) using site-specific vegetation coverage and topographic data extracted from Google Earth™ mapping service.

Model Setup - After collecting the necessary water-quality and site-specific data, model setup ensured water-quality and quantity data was apportioned correctly within the model framework (Table 1). The reach was segmented at 20, 100-meter sections spanning the 2 kilometer study area. Depth, velocity, bottom slope, and side slope point measurements were then interpolated between each reach segment. Each point and distributed source was assigned an average flow value.

Populating headwater data consisted of linear interpolation between available points to estimate hourly values. Point source information used either sonde information to produce a corresponding sine curve, or daily samples (average of two samples) to produce a constant daily concentration. Finally, any water-quality measurements from surface seeps or shallow groundwater observation wells were assigned to either a point (seep) or distributed inflow based on the evidence of groundwater inflow from the flow balance study.

Model Calibration - The methods described within the general calibration approach described previously were followed. However, additional steps included using the CBOD oxidation rates established from the WRF samples to convert the $cBOD_5$ samples to $CBOD_{ultimate}$. The most appropriate reaeration formula and a reach-wide SOD value were estimated for the site using stream metabolism methods estimated specifically using the Inverse Method. An SOD measurement using the chamber method (Hickey 1998) was later collected within the summer season of 2012 near Station 3 and used to evaluate the our approach and the performance. The remaining parameters were then optimized by the auto-calibration algorithm using the fitness statistic by combining the weighted normalized RMSE for each paired observation and prediction at Station 3 for all measured constituents (SI Table 3). Higher weights were assigned to overall indicator constituents such as DO min/max values and pH as well as key constituents such as inorganic phosphorus and nitrogen. While additional water-quality data were available at other stations and could have been incorporated in the fitness statistic for auto-calibration, we opted to rely on a downstream single station as the calibration target per the minimum requirements of the general calibration strategy. The additional data were used to gage the accuracy of the calibration based on limited data.

As described above, similar data collection methods were applied to six additional study reaches located throughout central to northern Utah (SI Figure 1) in order to develop and refine the generalized approach. Generally following the protocols outlined above, these models were populated and calibrated using data only from the most downstream sampling location (even when intermediate data were available). These model results provide further insight regarding the applicability and limitations of the outlined approach to a variety of point source influenced study reaches.

Results

Hourly averaged sonde data for Stations 1-3 and T1 (Figure 3) highlight the daily, longitudinal differences between stations. There are noticeable differences between DO at Station 1 versus those observed at Station 3. The point sources (Stations 2 and T1) differ significantly from each other, most notably between the diel signals, with the WRF (Station 2) experiencing minimal temperature diel variability and high variability for DO while the trends appear reversed for Station T1. Further, the differences of specific conductance between the main channel of Silver Creek and the main tributary would indicate that the source waters are distinct. Also shown are the chlorophyll *a* values (plotted on a log scale) which reflect some reasonable average values, however, the variability within the sonde measurements (as shown by the box plots) illustrates potential drawbacks of relying solely on optical in-situ measurements. Finally, the chemistry and nutrient data results were averaged by station and constituent (SI Table 4). Comparing values between Station 1 and 3 illustrate the effect Stations 2 and T1 have on the downstream concentrations. Also evident is the trade-off between *in situ* sonde measurements of chlorophyll *a* (diel response, larger variability, Figure 3) and the results from the laboratory analysis (fewer measurements, less variability and much lower values). The flow records taken from 8/18 to 8/30/2011 were averaged together by station and are compared with the net gains and losses results (bar chart) as shown in SI Figure 2. Shown along with the mean daily flow record is the number of points available to calculate the daily average (SI Figure 2b). Even with the low number of records generally available and the uncertainty in measurements, an estimation of the net water balance was necessary to ensure the correct volumes of water were represented in the model. To estimate net gains/losses, the measured mean daily flow value of an upstream station was subtracted from the nearest downstream station (SI Figure 2a). Surface seeps were independently estimated and assigned within the model as distributed sources centered near their surface location which combined represented 0.015 m^3/s or 13% of total stream flow.

Information required from the nutrient, chemistry, and sonde samples illustrate the differences in concentrations between longitudinal stations over a short (two day) period (SI Figure 3). Due to the high variability of the chlorophyll *a* data from the sondes, only results from the laboratory analysis were used for model setup. Further, since specific conductance and pH were measured using both methods (sonde and laboratory analysis) and both types of measurements compared reasonably well, all data were averaged together. All other constituents were summarized as average hourly or daily (in cases of limited samples) for the headwater, point sources, and other intermediate stations. Finally, any samples reported as below MDL were set to 0.5 MDL due to limited sample numbers.

Comparing predicted k_a values to those estimated using the Inverse Method indicated that either the USGS (channel-control/pool-riffle) or the Tsivoglou-Neal methods of reaeration align sufficiently to the data (SI Figure 4). Stream metabolism results (SI Table 5) from Eq. 1 produced GPP and ER values to estimate SOD ranging from 2.1 to 8.6 g $O_2/m^2/d$ with an average value of 5 g $O_2/m^2/d$. Since this average estimate represents the maximum SOD value possible, a slightly lower value of 3 g $O_2/m^2/d$ was prescribed as a conservative starting point for SOD along the entire reach. Due to the number and variability of point estimates, varying SOD by model reach could also be appropriate.

The overall calibration procedure produced reasonable results between observed and predicted values for many of the critical constituents required for accurate representation of stream water quality (Figure 4). Some constituents that matched observations well include inorganic and organic phosphorus as well as ammonium concentrations. Some areas of concern include temperature predictions which miss the mean and minimum observed values at some stations. Some possible explanations could be due to complex groundwater influences or topography and channel incision having a greater effect on shading than captured within the model. The average DO predictions match well, but the minimum and maximum values do not capture the observed diel swings in the upper half of the reach. The most significant source of nitrogen loading is from the WRF in the form of nitrate with values two orders of magnitude less for ammonium and one order less for organic nitrogen. The observed organic nitrogen values at Station 3 were well above the predicted concentrations, likely a legacy of using the differencing method for to derive its values. Predicted bottom algae concentrations were reasonable based on observations during September of 2011 where values ranged from 99 mgA/m^2 for a 70 m^2 section near Station 1, 150 mgA/m^2 for a 120 m^2 section near Station I2, and 226 mgA/m^2 for a 105 m^2 section near Station 3.

Overall, given the minimal amount of data used for model calibration (Station 3 water-quality), it appears that the model represents the observed conditions reasonably well, the exceptions are poor estimates of organic nitrogen, temperature, and DO at the intermediate stations. Including the data from these intermediate stations in the calibration likely would have improved the model calibration.

The six additional study sites throughout Utah (SI Figure 1) all had site specific conditions that influenced sampling, model setup, and/or model calibration (as described within the SI). Regardless, model performance was generally quite good for many key water quality constituents (SI Figure 5-10). It is clear, however, that additional information is necessary to corroborate predictions and that site specific issues influencing data and/or model predictions need to be investigated further. Neilson et al. (2012) provides generalized conclusions and recommendations from this effort.

Discussion

The generalized and basic data collection approach presented within this paper outlines the fundamental data types necessary to set up and conduct a preliminary calibration of the one-dimensional in-stream water-quality model QUAL2Kw assuming steady flow conditions. Since the approach was designed to be applied in the context of addressing various surface water-quality management objectives (e.g., WLAs, TMDLs, NNC), the need to adapt the approach to the case-specific requirements became apparent. The application of this guidance, presented in the context of a detailed case study and prior application to 6 other study areas during development of the generalized approach (see SI), illustrated the utility of both required and potential supplementary data to achieve acceptable predictions. Further, it highlighted the shortcomings of trying to develop low cost, minimalist data collection methods to support model setup and calibration.

Some of the site specific considerations for applying this approach included spatial and temporal sampling needs. Given sufficient background information about a site, some guidance suggests the selection of sampling locations along the reach be placed near the minimum of the DO sag since this will be the area where water-quality standards are likely to be violated (US EPA 1986). In the Silver Creek case study, we adequately captured the effects of the DO sag and nearly captured the downstream location where DO concentrations had returned to the upstream (pre-WRF) concentrations. Unfortunately, data from intermediate sampling locations did not capture this maximum sag in Silver Creek, but were still found useful in identifying locations where the model was obviously not capturing all important processes (Figure 4).

Similar to understanding station locations, bottom algae concentrations are a key factor in many shallow streams (Flynn and Suplee 2011) including the Silver Creek case study. However, there are still limitations when sampling one or two portions of a stream and generalizing the result to an entire study reach. Some methods have been developed which address proper sampling protocols and extrapolating results (CEN 2003), although inevitably, the underlying basis for small sample sizes generally involves time and cost constraints. Additional guidance is needed to derive reliable observations of reach-integrated bottom algae concentrations, to characterize filamentous algae and macrophytes, and how these can be incorporated into water-quality modeling efforts.

Beyond spatial considerations and based on the variability of flow and water-quality, sampling frequency requires more attention for most data types. Despite a low-flow, steady-state assumption, some observed within-day variability can be as large as changes seen on an annual timeframe (Nimick et al. 2011). This is possibly the case for Station 2, TN and TP (SI Figure 3). However, if sufficient diel information of each constituent for a specific site is gathered with the goal to identify a reduced sampling strategy without sacrificing the true signal of the data, the minimal temporal resolution necessary can be determined using spectral analysis with the selection of a sampling interval corresponding to the Nyquist frequency (US EPA 1982). Unfortunately, this data intensive exercise is not feasible for the majority of water-quality assessments. Alternatively, diel variations can be anticipated from previous efforts and daily sampling can be focused on times representing mean concentrations (Harrison et al. 2005; Nimick et al. 2011). The caveat to this is when the variability and magnitude of loading is more significant than the background variability and flow of the stream. In this case, a more extensive survey of water-quality will be required at the loading sources, particularly for unnatural loading signals (independent of photoperiod) commonly seen at a WRF effluent.

In this case study, a daily sampling strategy aimed at capturing anticipated minimum and maximum values appeared to be adequate for many of the necessary constituents. However, complications associated with small sample sizes were compounded when other constituents were estimated (e.g., organic nitrogen, detritus) or when irregular loads from the WRF influenced concentrations of a particular constituent. Further, constituents measured below the detection limit caused a significant bias in model setup and calibration. Due to the limited number of samples taken, the selection of an appropriate method to handle censored values was restricted because they represented more than 25% of the sample size (Berthouex and Brown 2002). The selection of appropriate methods to handle these data became simplified and arbitrary.

Overall prediction uncertainty was created by temporal and spatial data restrictions, difficulty estimating mean daily flows due to highly variable WRF loads, and limited methods for quantifying groundwater exchanges in dynamic systems. These influences were most apparent in the in-stream temperature predictions (Figure 4), but likely influenced the predictions of other constituents. Another key factor was the short reach length and travel time that influenced the ability to identify appropriate parameters that may be key in accurate scenario generation or extended model applications. Further, the assignment of a sine curve to represent the diurnal variation of point sources appears to be an inadequate representation due to an irregular (non-sinusoidal) daily signal (Figure 3). In these cases, the latest version of QUAL2Kw allows for input of hourly values for all sources.

While model selection choices are generally dependent on the management questions and accuracy requirements, the Silver Creek case study provides an example of a situation when the assumption of steady flow with variable concentration may be inappropriate. The QUAL2Kw modeling framework is capable of simulating non-steady flow conditions, but the approach outlined here applies to steady flow conditions due to project budgets often not being sufficient to support data collection for dynamic modeling. Clearly a balance must be achieved between limited data and adequate representation of the key processes and signals, but expanding the data collection to support calibration of non-steady conditions may avoid problems in systems like Silver Creek that experience drastic changes in loading from day to day. Kannel (2011) mentions similar limitations and emphasizes that despite these challenges and depending on the system, the time and cost advantages of assuming steady flows may outweigh the additional cost associated with calibration of continuous simulation of non-steady flows. In an effort to develop simple methods to set two key parameters, k_a and SOD, using DO time series and open-water metabolism methods, it appears that the proposed approaches are reasonable for Silver Creek. The assumptions associated with ER minus GPP being equivalent to a maximum SOD will not be applicable to all systems, however, in some circumstances, the ability to take advantage of already existing data to complete these calculations appears acceptable. Since there is no clear consensus on which methods are most appropriate for measuring or estimating SOD (Viollier et al. 2003) due to temperature gradients (Otubu et al. 2006), velocity dynamics (Nakamura and Stefan 1994), and spatial heterogeneity (Mugler et al. 2012), we chose to test our estimates by comparing them with SOD measurements from *in situ* chambers deployed in 2012. The chambers produced values near 3 g $O_2/m^2/day$ while our values ranged from 2.1-8.6 g $O_2/m^2/day$ with an average of 5.2 g $O_2/m^2/day$ based on the ER/GPP differencing calculations. These similarities suggest that the differencing approach is a reasonable way to set or bound SOD values before or during model calibration.

Reaeration is notoriously difficult to estimate (Genereux and Hemond 1992) though direct measurement techniques using a tracer such as propane to estimate the gas exchange coefficient are superior to deriving k_a using physical characteristics of bottom slope, water depth, and stream velocity (McCutchan et al. 1998). Using stream metabolism methods in the context of k_a values is not new (Odum 1956), however, using these values at many locations to inform the selection of the internal model formula has shown to be potentially useful. Unfortunately, in the case of Silver Creek, two formulas were found to have the best RMSE values although they differed significantly from one another (one higher, one lower). More importantly, neither were a good fit to the data due to simplified hydraulics within the model. While a poor formula fit might inspire a modeler to set a reach-wide k_a value using the average of observed data points, this practice can be problematic for scenarios run under different flow conditions. An alternate approach could be to calibrate the model based on the observed k_a values and use the data to select the most appropriate formula for subsequent scenarios.

Most process-oriented models are under-determined, wherein there exists more parameters than state variables to define them (Reckhow and Chapra 1999). Although a sensitivity analysis can help to identify the key processes influencing the state variables, there is an obvious need to decrease the number of parameters and potentially come up with narrower ranges to confine auto-calibration estimates. The potential number of calibration parameters is extremely high in QUAL2Kw, but without more information regarding which parameters are unimportant, it is not clear which should be dropped from the auto-calibration. More effort is needed to identify the most sensitive parameters of a system, narrow the reasonable ranges of those parameters, and set those parameter rates according to appropriate site-specific conditions. It is also important to identify which outputs must be included in a fitness statistic since the objective function (i.e., fitness) guides the auto-calibration. Given what we know of receiving streams downstream of WRFs, SOD and bottom plant growth are primary factors in governing DO dynamics (Chapra 2008; Utley et al. 2008) and future efforts should focus on developing appropriate sampling or simplified modeling approaches to represent these processes.

This study presents a minimalist approach to model setup and calibration that is reproducible and applicable to a diverse set of water-quality modeling problems. Applying this approach to the Silver Creek case study resulted in reasonable model predictions that captured many of the dominant processes that affect DO. However, for most problems this approach would only provide an initial framework for preliminary data collection and that would be adapted as needed. Additionally, a sensitivity analysis, additional model calibration, and model corroboration or validation would have to occur before applying it to management decisions. Regardless, this basic data collection approach results in an judicious use of resources while assisting in identifying the key factors requiring additional investigation.

Conclusion

In this paper, we developed a general data collection methodology to support QUAL2Kw model setup, explored methods for estimating key model parameters, and addressed the effectiveness of these methods with a case study of an effluent-dominated stream system. To minimize data collection costs, we identified the nominal number of sampling locations and minimal required data types for a WRF-dominated system. In the context of a case study we illustrated the utility of collecting grab samples over a two day period with one sample collected at dawn and the other at dusk. These data were supplemented with *in situ* sonde information to capture the daily variability of other constituents governed by the photoperiod. We found that this basic approach provided adequate information for model setup and calibration and reasonable predictions for the Silver Creek case study. However, we recognize that other sampling frequencies may be necessary for other study sites and objectives.

Some challenges were identified as we translated the collected data into model setup including estimation of correct flow rates and volumes, designation of appropriate loading values due to variable point source loads, and determination of appropriate calibration/fitness endpoints. The identification of key parameters (k_a and SOD) using data collected within the proposed methodology provided a means to decrease model uncertainty by reducing model parameters being calibrated. Future work to reduce the ranges of model calibration parameters, identification of sensitive parameters, and/or development of additional methods to set more model parameters based on site-specific conditions will help to increase confidence in model predictions. While this approach has merit as a starting point for WLAs, TMDLs, and in helping develop nutrient criteria, it should not be used as a "one-size-fits-all" strategy, but rather incorporated in an adaptive management strategy to guide more appropriate site-specific data collection schemes that will facilitate predictions needed to address management objectives.

References

Barnwell, T. O., L. C. Brown, et al. (2004). "Importance of field data in stream water quality modeling using QUAL2E-UNCAS." Journal of Environmental Engineering **130**: 643-647.

Bartholow, J. M. (1989). Stream temperature investigations: field and analytic methods. Instream Flow Information Paper U. S. F. a. W. Service. **13** 139.

Berthouex, P. M., and Brown, L. C. (2002). *Statistics for Environmental Engineers*. CRC Press, Boca Raton, FL.

Bischoff, J., Massaro, P., and Strom, J. (2010). *Jewitts Creek dissolved oxygen TMDL*. Wenck Associates, Inc., Maple Plain, MN, 23.

Boyacioglu, H., and Alpaslan, M. (2008). "Total maximum daily load (TMDL) based sustainable basin growth and management strategy." *Environmental Monitoring and Assessment*, 146(1), 411–421.

Brown, L. C., and Barnwell, T. O. J. (1987). *The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS*.

Carroll, J., O'Neal, S., and Golding, S. (2006). *Wenatchee river basin dissolved oxygen, pH, and phosphorus total maximum daily load study*. Washington State Department of Ecology, Olympia, 154.

CEN. (2003). *Water quality - guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers*. Comité European de Normalisation, Geneva.

Chapra, S. C. (2003). "Engineering water quality models and TMDLs." *Water Resources Planning and Management*, 129(4), 247–256.

Chapra, S. C. (2008). *Surface Water Quality Modeling*. Waveland Press Inc., Long Grove, IL.

Chapra, S. C., Pelletier, G. J., and Tao, H. (2006). *Qual2K: A modeling framework for simulating river and stream water quality*. Tuft University, Medford, MA.

Chapra, S. C., and Di Toro, D. M. (1991). "Delta Method For Estimating Primary Production, Respiration, And Reaeration In Streams." *Journal of Environmental Engineering*, 117(5), 640–655.

Charbonneau, P., and Knapp, B. (1995). *A user's guide to PIKAIA 1.0*. National Center for Atmospheric Research, Boulder, CO.

Chen, Y. D. (1996). *HSPF SHADE*. University of Georgia, Athens, GA.

Cirpka, O. A., Fienen, M. N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R., and Kitanidis, P. K. (2007). "Analyzing bank filtration by deconvoluting time series of electric conductivity." *Ground Water*, 45(3), 318–328.

Covino, T. P., and McGlynn, B. L. (2007). “Stream gains and losses across a mountain-to-valley transition: impacts on watershed hydrology and stream water chemistry.” *Water Resources Research*, 43(10), 1–14.

Di Toro, D. M. (2001). Sediment Flux Modeling. . New York, NY., Wiley-Interscience.

Di Toro, D. M. and J. F. Fitzpatrick (1993). Chesapeake Bay sediment flux model. Tech. . W. E. S. U.S. Army Corps of Engineers. Vicksburg, Mississippi. **Report EL-93-2:** 316.

Di Toro, D. M., P. R. Paquin, et al. (1991). "Sediment Oxygen Demand Model: Methane and Ammonia Oxidation." Journal of Environmental Engineering **116**(5): 945-986.

Dunnette, D. A. (1980). “Sampling frequency optimization using a water quality index.” *Water Pollution Control Federation*, 52(11), 2807–2811.

Ebel, B. A., and Loague, K. (2006). “Physics-based hydrologic-response simulation: seeing through the fog of equifinality.” *Hydrological Processes*, 20, 2887–2900.

Ecology, and Washington State Department of Ecology. (2003). *Shade.xls - a tool for estimating shade from riparian vegetation.*

Ensign, S. H., and Doyle, M. W. (2006). “Nutrient spiraling in streams and river networks.” *Journal of Geophysical Research*, 111, 1–13.

Facchi, A., Gandolfi, C., and Whelan, M. J. (2007). “A comparison of river water quality sampling methodologies under highly variable load conditions.” *Chemosphere*, 66(4), 746–756.

Flynn, K., and Suplee, M. (2011). *Using a computer water quality model to derive numeric nutrient criteria: Lower Yellowstone River*. Montana Department of Environmental Quality, Helena, MT.

Fogle, A. W., Taraba, J. L., and Dinger, J. S. (2003). “Mass load estimation errors utilizing grab sampling strategies in a karst watershed.” *Journal of the American Water Resources Association*, 39(6), 1361–1372.

Gammons, C. H., Babcock, J. N., Parker, S. R., and Poulson, S. R. (2011). “Diel cycling and stable isotopes of dissolved oxygen, dissolved inorganic carbon, and nitrogenous species in a stream receiving treated municipal sewage.” *Chemical Geology*, 283(1), 44–55.

Genereux, D. P., and Hemond, H. F. (1992). “Determination of gas exchange rate constants for a small stream on Walker Branch Watershed, Tennessee.” *Water Resources Research*, 28(9), 2365–2374.

Grace, M., and Imberger, S. (2006). *Stream metabolism: performing and interpreting measurements.* Water Studies Centre Monash University, Murray Darling Basin Commission and New South Wales Department of Environment and Climate Change, 204.

Guadagnini, A., and Neuman, S. P. (1999). “Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomnly nonuniform domains.” *Water Resources Research*, 35(10), 2999–3018.

Gunderson, L., and Klang, J. (2004). *Lower Minnesota River dissolved oxgygen TMDL*. Minnesota Pollution Control Agency.

Harrison, J. A., Matson, P. A., and Fendorf, S. E. (2005). "Effects of a diel cycle on nitrogen transformations and greenhouse gas emissions in a eutrophied subtropical stream." *Aquatic Science*, *67*, 308–315.

Harvey, J. W., Wagner, B. J., and Bencala, K. E. (1996). "Evaluating the reliability of the stream tracer approach to characterize stream-subsurface exchange." *Water Resources Research*, 32(8), 2441–2451.

Hattermann, F. F., Willems, P., and Kundzewicz, Z. W. (2010). "Modelling - a primer for practioners." *Water Framework Directive: Model Supported Implementation*, F. F. Hattermann and Z. W. Kundzewicz, eds., IWA Publishing, London, UK, 55–86.

Hazelton, C. (1998). "Variations between continuous and spot-sampling techniques in monitoring a change in river-water quality." *Water and Environment Journal*, 12(2), 124–129.

Henderson-Sellers, B., and Henderson-Sellers, A. (1996). "Sensitivity evaluation of environmental models using fractional factorial experimentation." *Ecological Modelling*, 86(2-3), 291–295.

Henjum, Mi. B., Hozalski, R. M., and Wennen, C. R. (2010). "A comparison of total maximum daily load (TMDL) calculations in urban streams using near real-time and periodic sampling data." *Journal of Environmental Monitoring*, 12(1), 234–241.

Hester, E. T., and Doyle, M. W. (2011). "Human impacts to river temperature and their effects on biological processes: a quantitative synthesis." *Journal of the American Water Resources Association*, *47*(3), 571–587.

Hickey, C. W. (1998). "Benthic chamber for use in rivers: testing against oxygen mass balances." *Journal of Environmental Engineering*, 114, 828–845.

Holtgrieve, G. W., Schindler, D. E., Branch, T. A., and A'mar, Z. T. (2010). "Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics." *Limnology and Oceanography*, *55*(3), 1047–1063.

Jones, J. B., Schade, J. D., Fisher, S. G., and Grimm, N. B. (1997). "Organic Matter Dynamics in Sycamore Creek, a Desert Stream in Arizona, USA." *Journal of the North American Benthological Society*, 16(1), 78–82.

Kannel, P. R., Kanel, S. R., Lee, S., Lee, Y.-S., and Gan, T. Y. (2011). "A review of public domain water quality models for simulating dissolved oxygen in rivers and streams." *Environmental Monitoring Assessment*, 16, 183–204.

Lettenmaier, D. P., Hooper, E. R., Wagoner, C., and Faris, K. B. (1991). "Trends in stream quality in the continental United States, 1978-1987." *Water Resources Research*, 24(3), 327–339.

McBride, G. B., and Chapra, S. C. (2005). "Rapid calculation of oxygen in streams: approximate delta method." *Journal of Environmental Engineering*, 131(3), 336–342.

McCutchan, J. H., Lewis, W. M., and Saunders, J. F. (1998). "Uncertainty in the estimation of stream metabolism from open-channel oxygen concentrations." *Journal of the North American Benthological Society*, 17(2), 155–164.

Mugler, C., Rabouille, C., Bombled, B., and Montarnal, P. (2012). “Impact of spatial heterogeneities on oxygen consumption in sediments: experimental observations and 2D numerical modeling.” *Journal of Geochemical Exploration*, 112, 76–83.

Nakamura, Y., and Stefan, H. G. (1994). “Effect of flow velocity on sediment oxygen demand: A theory.” *Journal of Environmental Engineering*, 120, 996–1016.

National Research Council. (2001). *Assessing the TMDL Approach to Water Quality Management*. The National Academies Press, Washington, DC.

National Research Council. (2007). *Models in environmental regulatory decision making*. National Research Council, Washington, DC.

Neilson, B. T., and Chapra, S. C. (2003). “Integration of water quality monitoring and modeling for TMDL development.” Water Resources Impact, 5(1), 9–11.

Neilson, B.T., Hobson, A.J., von Stackelberg, N., Ostermiller, J. (2012) "Using Qual2K Modeling to Support Nutrient Criteria Development and Wasteload Analyses in Utah." Final Project Report. Utah Department of Environmental Quality. Division of Water Quality. September 2012.

Nimick, D. A., Gammons, C. H., and Parker, S. R. (2011). “Diel biogeochemical processes and their effect on the aqueous chemistry of streams: A review.” *Chemical Geology*, 283(1), 3–17.

Odum, H. T. (1956). “Primary production in flowing waters.” *Limnology and Oceanography*, 1(2), 102–117.

Orlob, G. T. (1992). “Water-quality modeling for decision making.” *Journal of Water Resources Management and Planning*, 118, 295–307.

Ort, C., Lawrence, M. G., Reungoat, J., and Mueller, J. F. (2010). “Sampling for PPCPs in wastewater systems: comparison of different sampling modes and optimization strategies.” *Environmental Science & Technology*, 44(16), 6289–6296.

Otubu, J. E., Hunter, J. V., Francisco, K. L., and Uchrin, C. G. (2006). “Temperature effects on tubificid worms and their relation to sediment oxygen demand.” *Journal of Environmental Science and Health, Part A*, 41(8), 1607–1613.

Pelletier, G. J., and Chapra, S. C. (2008). *QUAL2Kw theory and documentation*. Washington State Department of Ecology, Olympia, WA.

Pelletier, G. J., Chapra, S. C., and Tao, H. (2006). “QUAL2Kw - A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration.” *Environmental Modelling & Software*, 21(3), 419–425.

Reckhow, K. H., and Chapra, S. C. (1999). “Modeling excessive nutrient loading in the environment.” *Environmental Pollution*, 100, 197–207.

Ruehl, C., Fisher, A. T., Hatch, C., Huertos, M. L., Stemler, G., and Shennan, C. (2006). “Differential gauging and tracer tests resolve seepage fluxes in a strongly-losing stream.” *Journal of Hydrology*, (330), 235–248.

Scholten, H., and Refsgaard, J. C. (2010). "Quality assurance in model-based water management." *Modeling Aspects of Water Framework Directive Implementation*, P. A. Vanrolleghem, ed., IWA Publishing, London.

Stahl, J. K., and Smith, R. L. (2002). *TMDL analysis for Limekiln Brook, Danbury, Connnecticut*. Department of Environmental Protection, Hartford, CT.

Turner, D. F., Pelletier, G. J., and Kasper, B. (2009). "Dissolved Oxygen and pH Modeling of a Periphyton Dominated, Nutrient Enriched River." *Journal of Environmental Engineering*, 135(8), 645–652.

UDEQ. (2000). *TMDL for East Canyon Creek*. Utah Department of Environmental Quality, Salt Lake City, UT.

UDEQ. (2012a). *Field data collection for QUAL2Kw model build and calibration, standard operating procedures, Version 1.0*. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, UT.

UDEQ. (2012b). *Utah wasteload analysis procedure*. Utah Department of Environmental Quality, Division of Water Quality, Salt Lake City, UT, 1–23.

US EPA. (1982). *Handbook for sampling and sample preservation of water and wastewater*. United States Environmental Protection Agency, Washington, DC.

US EPA. (1986). *Stream sampling for waste load allocation applications*. United States Environmental Protection Agency, Washington DC.

US EPA. (1995). *Technical guidance manual for for developing total maximum daily loads*. United States Environmental Protection Agency, Washington, DC.

US EPA. (1997). "Streams and Rivers." *Technical guidance manual for performing waste load allocations*, Streams and Rivers, United States Environmental Protection Agency, Washington, DC.

US EPA. (2000). *Nutrient criteria technical guidance manual: rivers and streams*. United States Environmental Protection Agency, Washington, DC.

US EPA. (2002a). *TMDL for total meurcury in fish tissue residue*. United States Environmental Protection Agency: Region 4, Lake Oconee, GA.

US EPA. (2002b). *Guidance on choosing a sampling design for environmental data collection*. United States Environmental Protection Agency, Washington, DC.

US EPA. (2009). *National Rivers and Streams Assessment Field Operations Manual*. United States Envrionmental Protetion Agency, Washington, DC.

Utley, B. C., Vellidis, G., Lowrance, R., and Smith, M. C. (2008). "Factors affecting sediment oxygen demand dynamics in blackwater streams of Georgia's coastal plain." *Journal of the American Water Resources Association*, 44(3), 751–753.

Viollier, E., Rabouille, C., Apitz, S. E., Breuer, E., Chaillou, G., Dedieu, K., Furukawa, Y., Grenz, C., Hall, P., Janssen, F., Morford, J. L., Poggiale, J. C., Roberts, S., Shimmield, T., Taillefert, M., Tengberg, A., Wenzhofer, F., and Witte, U. (2003). "Benthic biogeochemistry: state of the art technologies and guidelines for the future of in situ survey." *Journal of Experimental Marine Biology and Ecology*, 285, 5–31.

Vogt, T., Hoehn, E., Schneider, P., Freund, A., Schirmer, M., and Cirpka, O. A. (2010). "Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream." *Advances in Water Resources*, 33(11), 1296–1308.

von Stackelberg, N.O. and Neilson, B.T. (2014). "A collaborative approach to calibration of a riverine water quality model." Journal of Water Resources Planning and Management, 140(3):393-405, doi: 10.1061/(ASCE)WR.1943-5452.0000332.

Whitehead, J., and Judd, H. (2004). *Silver Creek total maximum daily load for dissolved zinc and cadmium*. Division of Water Quality, Salt Lake City.

Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., and Wade, A. J. (2009). "A review of the potential impacts of climate change on surface water quality." *Hydrological Sciences*, 54(1), 101–123.

Young, R., Townsend, C., and Matthaei, C. (2004). *Functional indicators of river ecosystem health - an interim guide for use in New Zealand*. Cawthron Institute, Nelson, New Zealand.

Zhang, J., and Zhang, C. (2012). "Sampling and sampling strategies for environmental analysis." *International Journal of Environmental Analytical Chemstry*, 92(4), 466–478.

FIGURE CAPTIONS

FIGURE 1. GENERALIZED DATA COLLECTION LOCATIONS ARE SHOWN WITH THE REQUIRED LOCATIONS OF FLOW MEASUREMENTS, MULTI-PARAMETER WATER QUALITY SONDES, AND CHEMISTRY AND NUTRIENT SAMPLES. HEADWATER (UPSTREAM BOUNDARY CONDITION) IS DESIGNATED BY STATION 1, THE PRIMARY POINT SOURCE, IS REPRESENTED BY STATION 2, TRIBUTARIES ARE DENOTED WITH T1, DIVERSIONS WITH D1, AND THE DOWNSTREAM CALIBRATION STATION IS SHOWN AS STATION 3.

Figure 2. The location of the study area includes a USGS station (river km 1.8) and the Silver Creek WRF (a). Also shown is the site schematic for the study reach (b). Major stations include (1) headwater at river km 2.0, (2) WRF point source at river km 1.9, (T1) a tributary to the stream at river km 1.2, and (3) a downstream calibration station at river km 0. Also shown are intermediate measurement stations denoted as USGS, I1, and I2 at river kilometers 1.5, 1.1, and 0.8, respectively. Visible waters flowing to the stream are denoted as surface seeps with S1, S2, and S3 at river kilometers 1.4, 1.0, and 0.3, respectively.

FIGURE 3. SONDE DATA COLLECTED AT STATION 1 (HEADWATER), STATION 2 (WRF), STATION T1 (TRIBUTARY), AND STATION 3 (CALIBRATION STATION) MEASURING TEMPERATURE, DO, PH, SPECIFIC CONDUCTANCE, AND CHLOROPHYLL A. THE FIRST COLUMN SHOWS EACH STATION PLACED IN ORDER LONGITUDINALLY ALONG THE REACH. THE SECOND COLUMN SHOWS THE DIFFERENCE FROM IN-STREAM STATIONS, THE HEADWATER AND THE DOWNSTREAM CALIBRATION STATION. THE THIRD COLUMN SHOWS THE MAJOR POINT SOURCES INFLUENCING THE STREAM REACH (WRF, T1). DIEL VALUES WERE AVERAGED HOURLY OVER TWO DAYS. NOTE THAT CHLOROPHYLL A VALUES WERE LOG_{10} TRANSFORMED DUE TO MANY EXTREME VALUES NEAR THE MAXIMUM DETECTION LIMIT.

Figure 4. Comparison of predicted versus measured data for a) flow, b) water temperature, c) DO, d) nitrate, e) ammonium, f) organic nitrogen, g) inorganic phosphorus, h) organic phosphorus, and i) bottom algae of Silver Creek (X axis is in river kilometers) for the Qual2Kw model calibration. The solid lines indicate model predictions, dashed lines are minimum and maximum predicted values, solid circles are average daily measurements and white circles are daily minimum and maximum observed concentrations.

TABLE 1. GENERAL INFORMATION REQUIRED FOR QUAL2KW MODEL SETUP.

QUAL2Kw Input	Information Required
Reach	Reach segmentation
	Hydraulic characteristics
	% suitable substrate
	Bottom algae % cover
	Sediment Oxygen Demand (SOD)
	Thermal properties
Initial Conditions	Constituent concentrations[1, 2]
Headwater Data	Average flow
	Hourly mean concentrations[1]
Weather Data	Hourly mean values[3]
Point Sources	Average flow
	Daily mean[1], Range/2, Time of Max
Distributed Sources	Average flow
	Daily mean concentrations[1]
Point and Distributed Abstractions	Average flow
Rates	Primarily set in calibration
	Literature informed ranges of parameters

[1]See SI Table 1for a list of constituents required by QUAL2Kw.
[2]Optional.
[3] The model is interpolating between the hourly values for each time step, therefore, the input data should ideally be an average of the instantaneous data on the hour. If using hourly averages, the averages should be centered on the hour.

TABLE 2. SITE CHARACTERIZATION DATA TYPES, PROCEDURES AND LOCATIONS.

Data Type	Procedure	Locations	Reasoning
Average Cross Sectional Velocity	Velocity cross-sectional profile obtained from velocity-area method of discharge measurements. Information from HEC-RAS modeling applications can also be extracted to supplement data collected.	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of velocity in different reaches to compare with the predicted velocities. This can be used with the depth and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Cross Sectional Depth	Average depth can be obtained from velocity-area method of discharge or independently estimated cross-sectional depth profiles.	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Provides observations of depths in different reaches to compare with the predicted depths. This can be used with the velocity and tracer information to ensure appropriate representation of the hydraulics and reasonable travel times.
Average Channel Bottom Width	Bottom width estimates are calculated using the formula: Top Width (m) $–\text{Depth}_{ave}$ (m) x $1/\tan[\text{radians}(°SS_{LEW})]$ $– \text{Depth}_{ave}$ x $1/\tan[\text{radians}$ $(°SS_{REW})]$.	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model Input. Top widths, side slopes, and bottom slopes are measured at consistent increments along the channel. From these data, bottom width estimates can be calculated using side slope, average depth, and top width values.
Channel Bottom Slope	Measured with a survey level or clinometers, protocols described within EMAP documentation[1].	Should estimate bottom slope from beginning to end of study reach at each station and/or when changes in bottom slope are observed.	Model Input. Bottom slope affects travel time and can be adjusted along with Manning's n for to achieve proper estimates.
Channel Side Slope	Measured with a clinometer or by visual inspection, protocols described within EMAP documentation[1].	Station 1, 3, and above and below any inflow or outflow. Additional locations along study reach would be beneficial.	Model input. Can be used to calculate bottom width from measured top widths.
Weather data	Onsite or nearest weather station.	Near study site would be most appropriate and 15-30 minute data are preferred, hourly estimates required.	Wind speed, air temperature, shortwave solar radiation, humidity/dew point temperature are all used within the model as forcing information.
Tracer Study	Inject tracer at Station 1 or 2 and measure response at Station 3. . Can also use HEC-RAS model if available.	Measure tracer response at Station 3, but additional locations along the study reach would be beneficial to capture heterogeneity and identify potentially significant groundwater sources.	Provides information regarding average travel time through system and can be used in calibration of hydraulic parameters (e.g., Manning's roughness coefficient).
Substrate type	Protocols described within EMAP documentation[1]	Information should be gathered at cross sections in sub-reaches that represent the variability in substrate types.	Provides a method to approximate the Manning's roughness coefficient and determine fraction of bottom substrate appropriate for bottom algae.

Shading	Estimated with shading model (e.g., SHADE[2]) to predict effective shade from topography and riparian vegetation	Information should be gathered at locations that represent the variability in shading.	Model input. If riparian or topographic shading drastically influences in-stream temperatures, estimates of the shading % for each hour of a day will be necessary to scale the incoming shortwave solar radiation.

[1](US EPA 2009) [2](Ecology and Washington State Department of Ecology 2003)

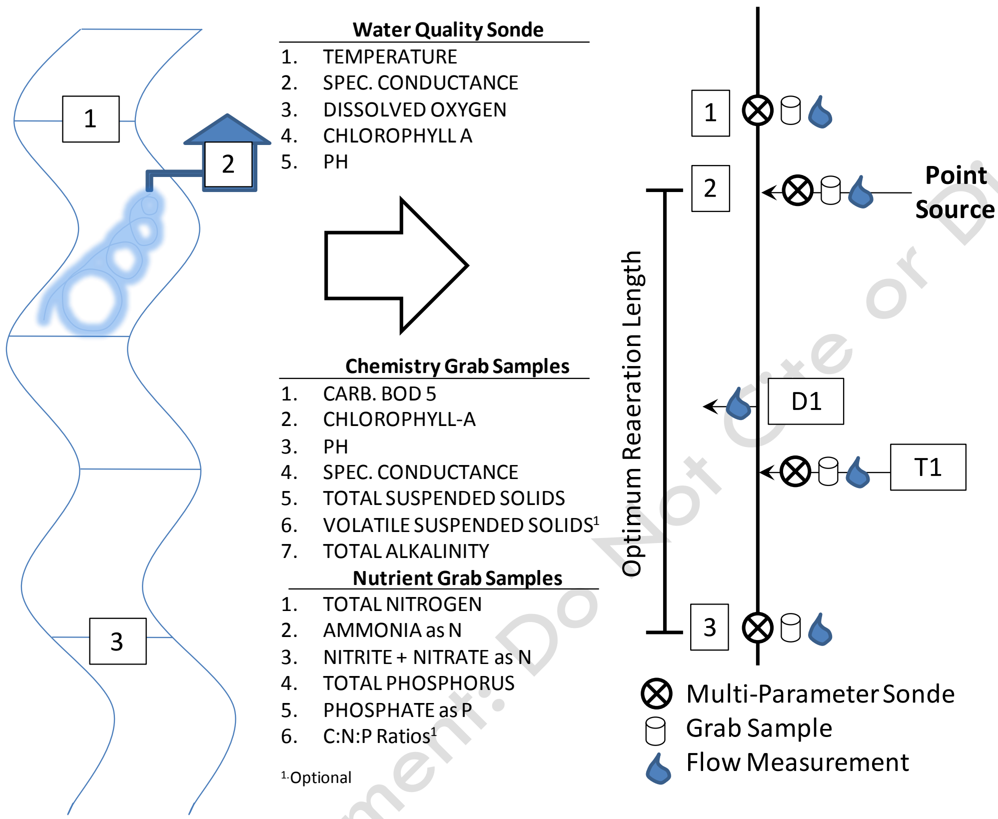


FIGURE 1. GENERALIZED DATA COLLECTION LOCATIONS ARE SHOWN WITH THE REQUIRED LOCATIONS OF FLOW MEASUREMENTS, MULTI-PARAMETER WATER QUALITY SONDES, AND CHEMISTRY AND NUTRIENT SAMPLES. HEADWATER (UPSTREAM BOUNDARY CONDITION) IS DESIGNATED BY STATION 1, THE PRIMARY POINT SOURCE IS REPRESENTED BY STATION 2, TRIBUTARIES ARE DENOTED WITH T1, DIVERSIONS WITH D1, AND THE DOWNSTREAM CALIBRATION STATION IS SHOWN AS STATION 3.

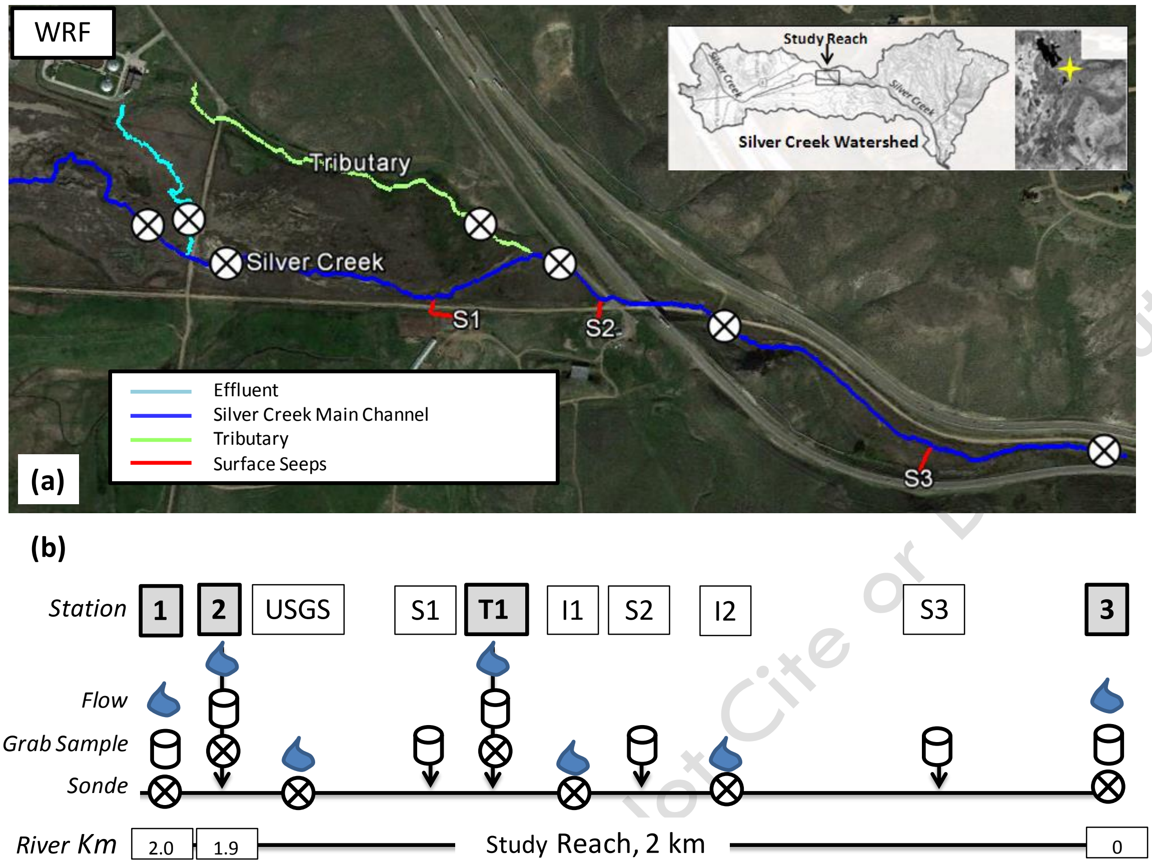


Figure 2. The location of the study area spans just upstream of the discharge of the Silver Creek WRF to Silver Creek and extends downstream approximately 2 km for a total travel time of 0.13 days (≈ 3 hours). Water quality is continuously monitored at a USGS station (river km 1.8) and the Silver Creek WRF (a). The aerial map is simplified into a site schematic for the study reach (b). Major stations include (1) headwater at river km 2.0, (2) WRF point source at river km 1.9, (T1) a tributary to the stream at river km 1.2, and (3) a downstream calibration station at river km 0. Also shown are intermediate measurement stations denoted as I1 and I2 at river kilometers 1.1 and 0.8, respectively. Visible waters flowing to the stream are denoted as surface seeps with S1, S2, and S3 at river kilometers 1.4, 1.0, and 0.3, respectively.

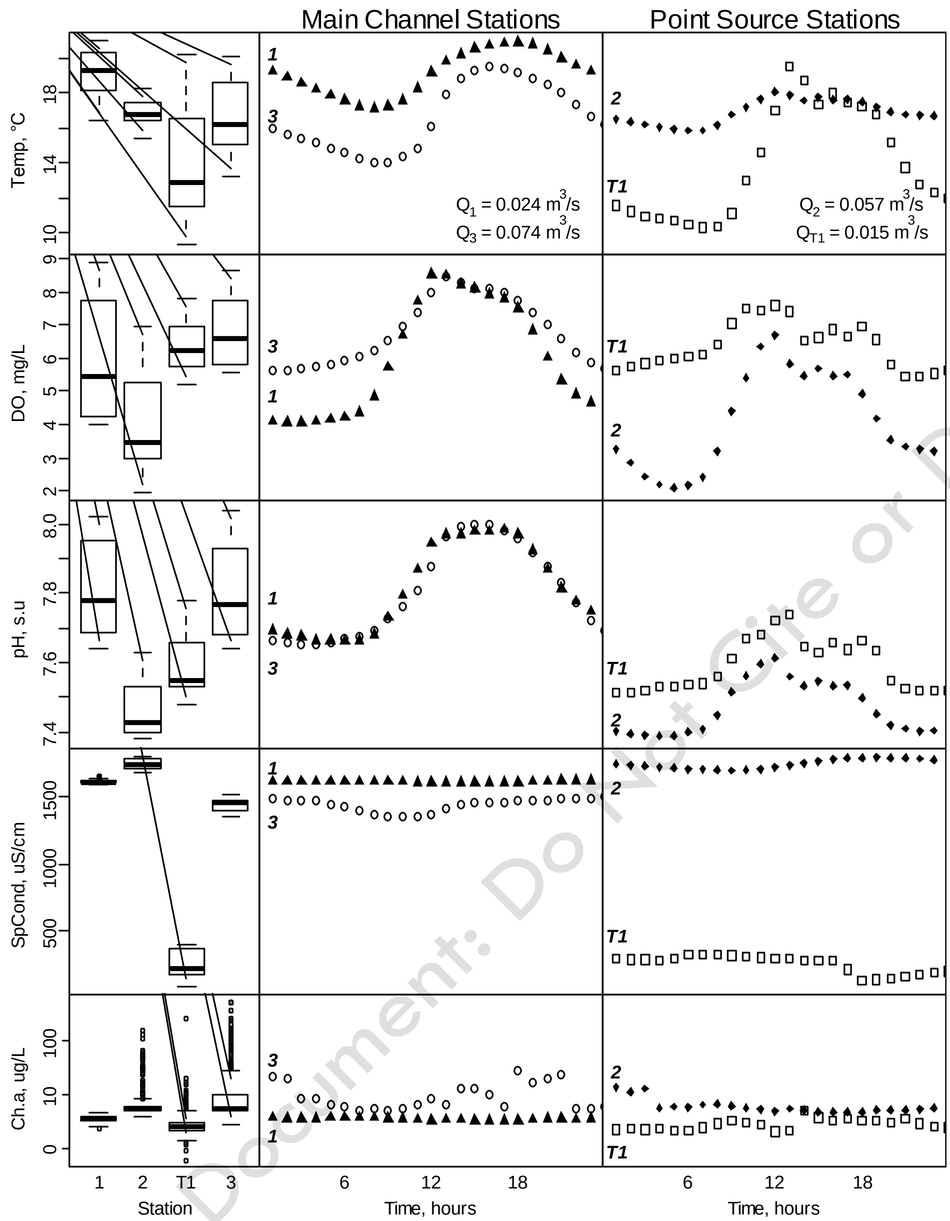


FIGURE 3. SONDE DATA COLLECTED AT STATION 1 (HEADWATER), STATION 2 (WRF), STATION T1 (TRIBUTARY), AND STATION 3 (CALIBRATION STATION) MEASURING TEMPERATURE, DO, PH, SPECIFIC CONDUCTANCE, AND CHLOROPHYLL A. THE FIRST COLUMN SHOWS EACH STATION PLACED IN ORDER LONGITUDINALLY ALONG THE REACH. THE SECOND COLUMN SHOWS THE DIFFERENCE FROM IN-STREAM STATIONS, THE HEADWATER AND THE DOWNSTREAM CALIBRATION STATION. THE THIRD COLUMN SHOWS THE MAJOR POINT SOURCES INFLUENCING THE STREAM REACH (WRF, T1). DIEL VALUES WERE AVERAGED HOURLY OVER TWO DAYS. NOTE THAT CHLOROPHYLL A VALUES WERE LOG$_{10}$ TRANSFORMED DUE TO MANY EXTREME VALUES NEAR THE MAXIMUM DETECTION LIMIT.

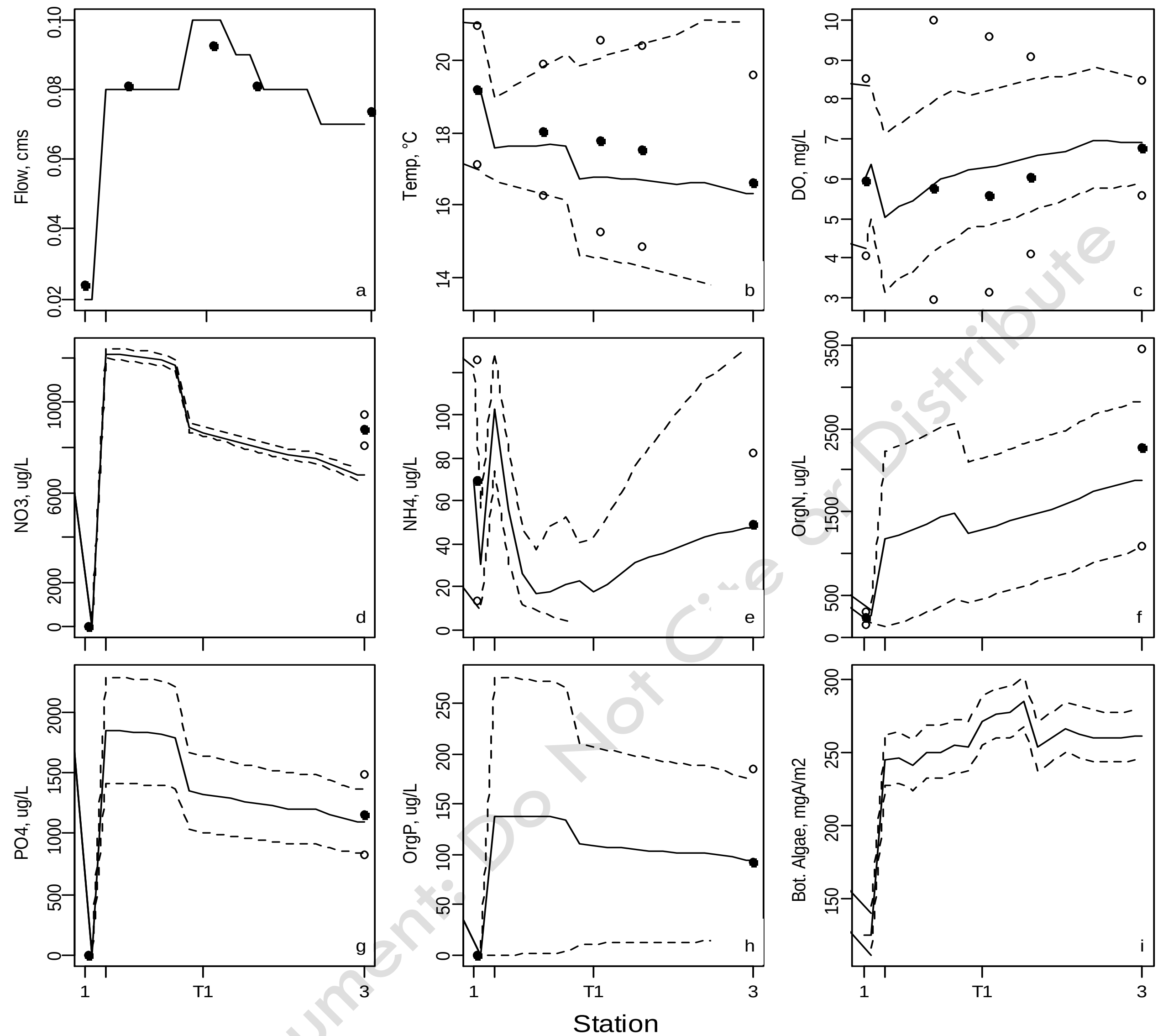


Figure 4. Comparison of predicted versus measured data for a) flow, b) water temperature, c) DO, d) nitrate, e) ammonium, f) organic nitrogen, g) inorganic phosphorus, h) organic phosphorus, and i) bottom algae of Silver Creek (X axis is in river kilometers) for the Qual2Kw model calibration. The solid lines indicate model predictions, dashed lines are minimum and maximum predicted values, solid circles are average daily measurements and white circles are daily minimum and maximum observed concentrations.

Sources of uncertainty in nutrient collection methods below a point source

Hayden Campbell, Undergraduate Researcher (Biology)
Michelle Baker (Biology), Beth Neilson (Civil and Environmental Engineering), and Susan Durham (Ecology Center Statistician).

14 February 2014

ABSTRACT

The goal of this study was to determine what aspects of sampling and sample storage could lead to uncertainty when taking samples below a point source. Sources of uncertainty studied were the locations where the samples were taken to assess if nutrients were adequately mixed within a cross-section, different filtration techniques, dilution errors, analytical uncertainty, and storage time. Bootstrapping analyses were used to determine whether mixing and dilution errors led to uncertainty, while one-way ANOVAs were used to evaluate filtration techniques and storage time. Sample spikes to determine percent recovery of nutrients and repeat sample analyses are routinely performed as part of the lab quality assurance/quality control plan (QA/QC), and are used here to evaluate analytical uncertainty. Comparison of coefficients of variation (COVs) of samples collected within a cross section at four locations, above, at, and below a point source, revealed that mixing of nutrients within a cross section appeared to be different at the different locations. The filtration devices analyzed were an electric pump and a manual syringe. These two devices gave statistically similar results in nitrate and soluble reactive phosphorus concentrations ($p>0.05$), but syringe-filtered samples had significantly higher ammonium concentrations ($p<0.05$). Dilution error was determined by comparing seven diluted samples with the original sample with which they were made. Dilutions proved to have the highest uncertainty relative to other treatments. The diluted samples were consistently higher than the original sample for all nutrients and were more variable than lab QA/QC duplicates for ammonium and soluble reactive phosphorus. Analytical uncertainty was found to be less than uncertainty associated with sample collection and storage except for unanticipated protocol failure. For this study, QA/QC data beyond 20% were considered fails, and the samples required reanalysis. In most cases the percent recovery of spiked samples and coefficients of variation of samples repeatedly analyzed were much less than 20%. However, ammonium and total nitrogen incurred the most failures. Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations ($p<0.05$), however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates.

INTRODUCTION

Nutrient samples are often collected below point sources, such as wastewater treatment plants, to ascertain if nutrient quantities exceed in-stream water quality standards. When analyzing these samples, it is imperative that the samples at the time of analysis are representative of the samples at the original time of sampling. In order to ensure that nutrient samples are reliable, this study tested five sources of uncertainty that had the potential to cause unreliable nutrient measurements. These sources included the location within a cross-section that samples were taken in order to assess whether inadequate mixing was occurring within the stream, different filtration techniques, dilution errors, analytical uncertainty, and storage time. Samples were collected in Silver Creek around the Silver Creek Water Reclamation Facility near Park City, Utah.

This research emerged due to recognized anomalies in previous nutrient sampling completed below point sources. Previously, when these analyses were completed, comparisons were made between total phosphorus and constituent phosphorus concentrations (e.g., soluble reactive phosphorus (SRP)). The amount of SRP measured was greater than the amount of total phosphorus measured. Additionally, similar anomalies were found when comparing constituent dissolved nitrogen concentrations and total nitrogen concentrations. This led to the recognition of sampling and/or analytical errors, but did not reveal the source of the error. This research was conducted in order to determine what potential sources of uncertainty could have led to these anomalies.

METHODS

Sampling Location

Four locations were chosen along Silver Creek where samples were collected to test whether mixing could cause different nutrient concentrations depending on where in the cross section the samples were collected. These locations included one above where the wastewater treatment plant effluent enters the stream (Above WWTP, approximately 13 meters), one in the wastewater effluent (Point Source), and two below where the wastewater treatment plant effluent enters the stream (Below (I) WWTP, approximately 103 meters, and Below (II) WWTP, approximately 717 meters) (Figures 1a and 1b). Seven 1,000 mL grab samples were collected at each of the four cross-sections. Each sample was collected at a different location within the cross section, both at varying distances across the cross section and at varying depths (Figure 2). From each 1,000 mL sample, two sub-samples were taken. One sub-sample was 120 mL that was not filtered, and was used to analyze for total nitrogen (TN) and total phosphorus (TP). The second sub-sample was filtered with a syringe and was analyzed for nitrite+nitrate-N, ammonium-N, and soluble reactive phosphorus (SRP). Samples were collected in the following order to prevent contamination: Below (II) WWTP, Below (I) WWTP, Point Source, Above WWTP.

The coefficients of variation were calculated for each cross section and compared to determine if they were statistically different. This was done in order to identify if the cross sections, whether above or below the wastewater treatment plant, had different mixing patterns. Three parametric tests were run to compare all four coefficients of variation, including the Modified Bennet's test, the Wald Test, and the Modified Miller Test (Jafari & Kazemi, 2013). These tests would determine statistical significance if the calculated p-value was less than 0.05. Since these tests compared all four sites together, another statistical test was required in order to determine which sites, between the four, actually were statistically different. A nonparametric bootstrap was used to compare the coefficients of variation between two individual sites to determine which sites yielded different mixing patterns.

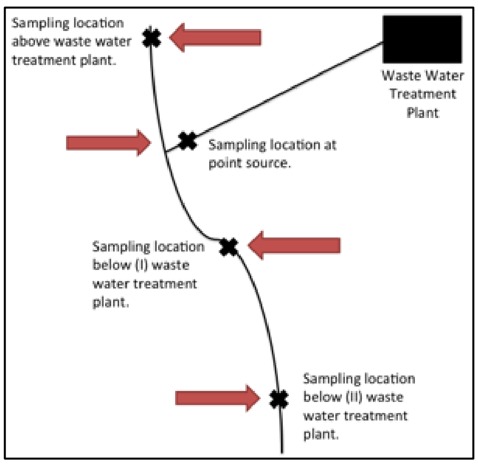


Figure 5a. Locations for testing inadequate mixing.

FIGURE 1B. GOOGLE EARTH IMAGE OF SAMPLING LOCATIONS

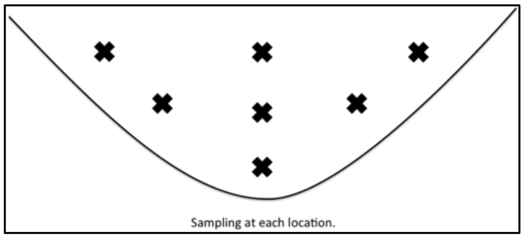


Figure 2. Approximate locations in cross section of stream where grab samples are obtained.

Filtration Techniques

The samples collected for analyzing whether differences in nutrient concentrations exist when filtering with a syringe or an electric pump were collected at the farthest location downstream (Below (II) WWTP) (Figure 3). A 5-gallon bucket was collected in the middle of the stream at this location. From this bucket, which was kept mixed with a hand mixer, 12 samples were collected using the electric geopump (Figure 4), and 12 samples were collecting using a manual syringe. Each of these samples was analyzed for ammonium, nitrate, and soluble reactive phosphorus (SRP). One-way ANOVAs were used to compare the average concentrations of each nutrient for both filtering methods.

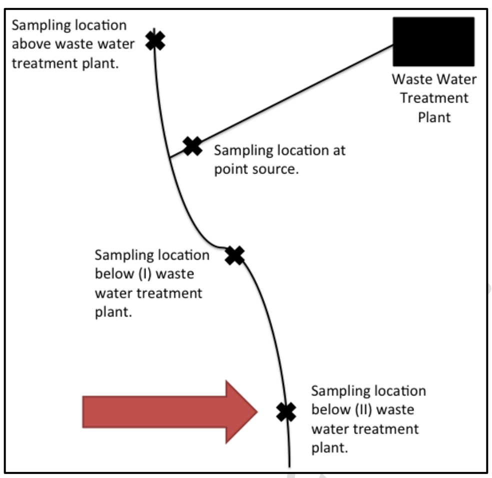


Figure 3. Location of sampling to contrast filtering methods.

FIGURE 4. ELECTRIC PUMP AND SYRINGES USED TO CONTRAST FILTERING METHODS.

Dilution Error

From another five-gallon bucket collected Below (II) WWTP, five 120 mL grab samples were taken and five samples were filtered with a syringe. These samples were repeat samples to determine how variable nutrient concentrations were when samples were collected from the same location. The unfiltered samples were analyzed for total nitrogen and total phosphorus, and the filtered samples were analyzed for ammonium, nitrate, and soluble reactive phosphorus. Two unfiltered samples and two filtered samples were used to make dilutions to determine how variable nutrient concentrations were when making dilutions. Five 1:100 dilutions were made for each of the four samples. The dilutions made on the unfiltered samples were analyzed for total nitrogen and total phosphorus, and the dilutions made on the filtered samples were analyzed for ammonium, nitrate, and soluble reactive phosphorus. These concentrations were compared to the original samples.

Analytical Uncertainty

Lab quality assurance and control (QA/QC) uses results from spiked samples, duplicates, and certified reference materials to assess lab analyses. For this study, QA/QC data beyond 20% were considered fails, and the samples required reanalysis. For each nutrient, the number of fails were tallied for duplicate samples and spiked samples to determine what nutrients incurred the most analytical failure.

Storage Time

Samples to assess the reliability of freezing were collected from Below (II) WWTP. A 4,000 mL grab sample was collected from the middle of the stream and transported back to the lab.

Twenty-eight samples were made from this 4,000 mL sample and placed in the freezer. Because these samples were unfiltered, they were analyzed for total nitrogen and total phosphorus. Seven samples were analyzed after one week of freezing, seven samples were analyzed after three weeks of freezing, seven samples were analyzed after six weeks of freezing, and seven samples were analyzed after twelve weeks of freezing. The average concentrations were compared using one-way ANOVAs to test for a statistical significance.

After the seven samples were analyzed after one week, they were placed back in the freezer and analyzed again at three weeks, six weeks, and twelve weeks of freezing. The samples analyzed after three weeks were also placed back in the freezer and analyzed again at six weeks and twelve weeks of freezing. The same was done for the samples analyzed after six weeks (i.e., they were placed back in the freezer after analysis and analyzed again at twelve weeks). This was done to determine whether multiple thawing and freezing events affected nutrient concentrations. One-way ANOVAs were used to compare the nutrient concentrations between re-freezing events.

RESULTS

Sampling Location

The average nutrient concentrations for the seven samples representing the cross section collected furthest downstream the wastewater treatment plant (Below (II) WWTP) were first compared with five duplicate samples (Figures 5-9). The five duplicate samples were taken from a bucket collected from the middle of the stream also at the furthest location downstream the wastewater treatment plant (Below (II) WWTP). Five filtered samples and five unfiltered samples were taken from the bucket. These samples were to represent samples that we would assume to be completely mixed. Because the five samples were collected from a mixed bucket, these samples were expected to be less variable. However, this was not observed for total nitrogen, nitrate, or soluble reactive phosphorus (Figures 6, 8, and 9). For these nutrients, the samples collected across the cross-section were less variable than the samples taken from the mixed bucket.

The Modified Bennet's test, the Wald test, and the Modified Miller test showed statistically significant differences between coefficients of variation for all nutrients between the four locations (Figure 1a), except for total nitrogen ($p>0.05$) (Figure 10). However, when pairwise comparisons using nonparametric bootstrap analyses were computed for each nutrient (Figure 11), a statistical significance was observed for total nitrogen between two pairs of locations. Statistical significance was also observed between coefficients of variation for the four other nutrients between at least one pair of locations, these p-values are highlighted in red in Figure 11. Despite these differences, no pattern was observed across all nutrients. However, the coefficients of variation appeared to be lowest for samples collected at the point source in all nutrients compared to the other three locations. The nonparametric bootstrap analyses did not give all the expected results, seen by the blue value in Figure 11. A statistical significance was expected for this value because of the difference between the coefficients of variation between the two locations. An outlier in the data probably caused the observed results.

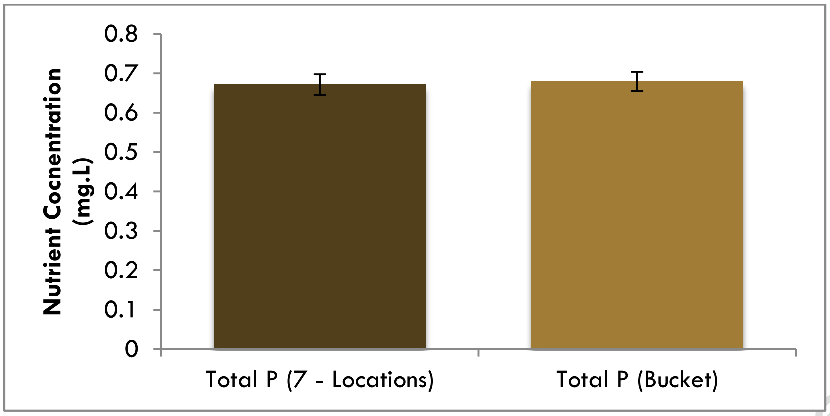


Figure 5. Comparison of total phosphorus concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

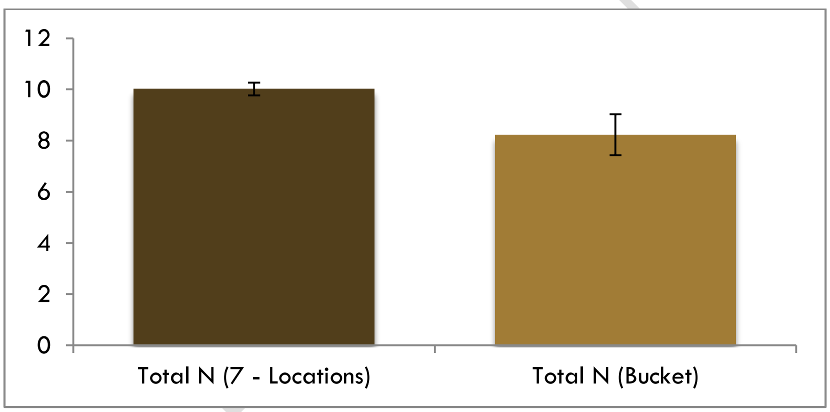


Figure 6. Comparison of total nitrogen concentrations between seven samples collected at various locations within the cross section, and five samples collected from a mixed bucket. The error bars represent standard error between samples.

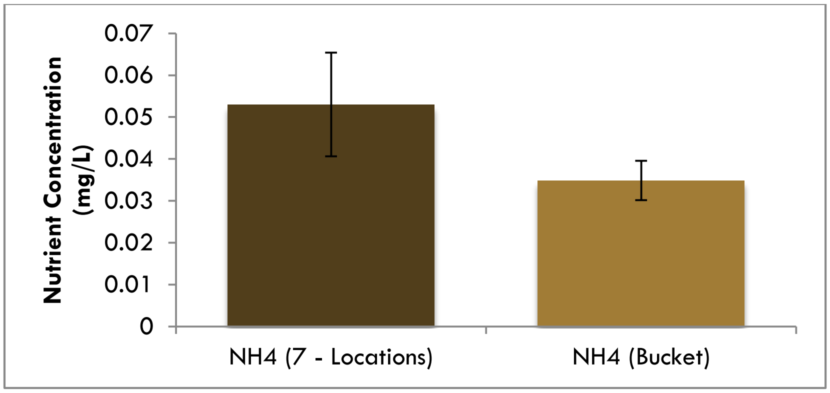


Figure 7. Comparison of ammonium concentrations between seven samples collected at various locations within the cross section, and five samples collected from mixed bucket. The error bars represent standard error between samples.

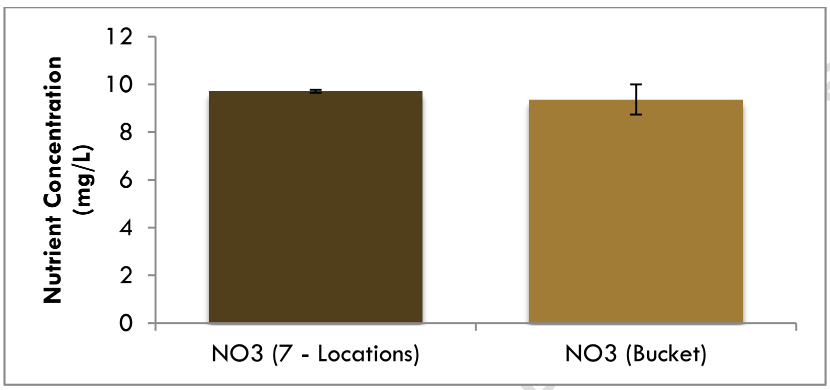


Figure 8. Comparison of nitrate concentrations between seven samples collected at various locations within the cross section, and five samples collected from mixed bucket. The error bars represent standard error between samples.

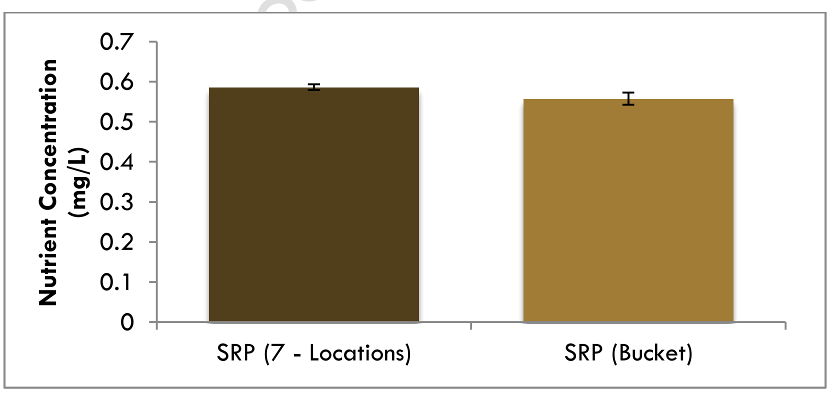


Figure 9. Comparison of soluble reactive phosphorus concentrations between seven samples collected at various locations within the cross section, and five samples collected from mixed bucket. The error bars represent standard error between samples.

TP:

	P-Value
Modified Bennet's test	9.538705e-11
Wald test	7.770074e-07
Modified Miller test	2.340471e-12

TN:

	P-Value
Modified Bennet's test	0.06184338
Wald test	0.06317188
Modified Miller test	0.06858057

NH4:

	P-Value
Modified Bennet's test	0.02990244
Wald test	0.0453102
Modified Miller test	0.04590821

NO3:

	P-Value
Modified Bennet's test	9.94E-35
Wald test	4.75E-09
Modified Miller test	4.27E-29

SRP:

	P-Value
Modified Bennet's test	9.25E-21
Wald test	1.43E-08
Modified Miller test	1.88E-24

Figure 10. P-values computed from the Modified Bennet's test, the Wald test, and the Modified Miller test.

TP:

CV	0.2236685	0.02277014	0.08744049	0.1020543
	Above	Point Source	Below I	Below II
Above	x	0.02339532	0.1143771	0.1785643
Point Source	x	x	0.01939612	0.02579484
Below I	x	x	x	0.6540692
Below II	x	x	x	x

TN:

CV	0.06476794	0.1677015	0.1814677	0.06576692
	Above	Point Source	Below I	Below II
Above	x	0.08638272	0.0689862	0.7940412
Point Source	x	x	0.8408318	0.01779644
Below I	x	x	x	0.0269946
Below II	x	x	x	x

NH4:

CV	0.2517278	0.1308364	0.3704142	0.6181216
	Above	Point Source	Below I	Below II
Above	x	0.06778644	0.3537293	0.0679864
Point Source	x	x	0.09338132	0.2623475
Below I	x	x	x	0.4027195
Below II	x	x	x	x

NO3:

CV	0.6896826	0.01786509	0.02435554	0.01606915
	Above	Point Source	Below I	Below II
Above	x	0.01219756	0.01339732	0.0109978
Point Source	x	x	0.1269746	0.4727055
Below I	x	x	x	0.07218556
Below II	x	x	x	x

SRP:

CV	0.3125886	0.01331167	0.03949525	0.03206434
	Above	Point Source	Below I	Below II
Above	x	0.01539692	0.01639672	0.01259748
Point Source	x	x	0.02859428	0.0319936
Below I	x	x	x	0.2733453
Below II	x	x	x	x

Figure 11. P-values computed from nonparametric bootstrap analyses comparing two locations. Note that x is included in the table where no comparison is needed (i.e., the CV for above does not need to be compared to itself) or where value is already provided in the table and red values indicate a statistical significance.

Filtration Techniques

After conducting one-way ANOVAs using the average concentration for the twelve samples filtered with the geopump and the twelve samples filtered with a manual syringe (Figures 12-14), the only significant differences between nutrient concentrations were found in ammonium concentrations (Figure 12). The p-value determined for ammonium was 0.049, while the p-values for nitrate and SRP were 0.445 and 0.286, respectively. The ammonium concentrations were consistently higher in samples filtered with the syringe than with the pump. The difference between the average ammonium concentrations filtered with the syringe and with the pump was also greater than the method detection limit, verifying the statistical significance.

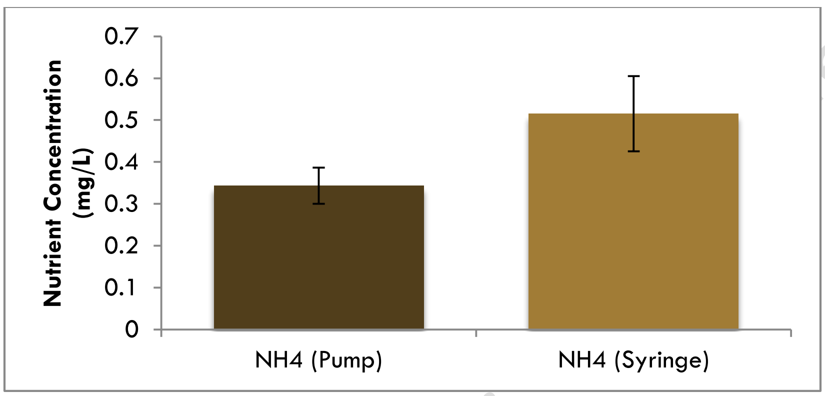


Figure 12. Comparison of ammonium concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

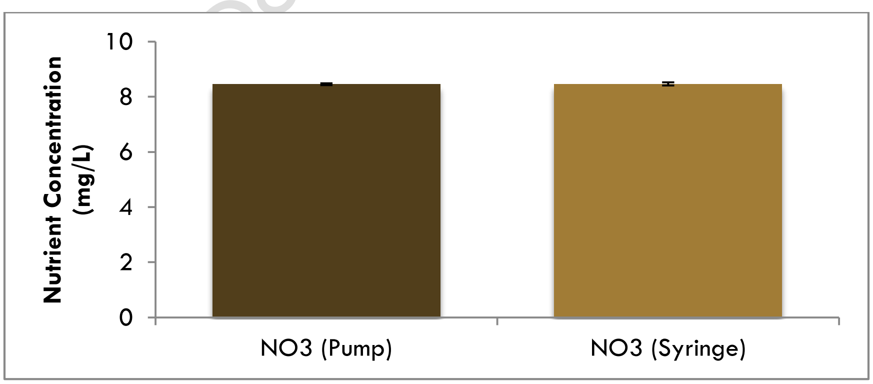


Figure 13. Comparison of nitrate concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

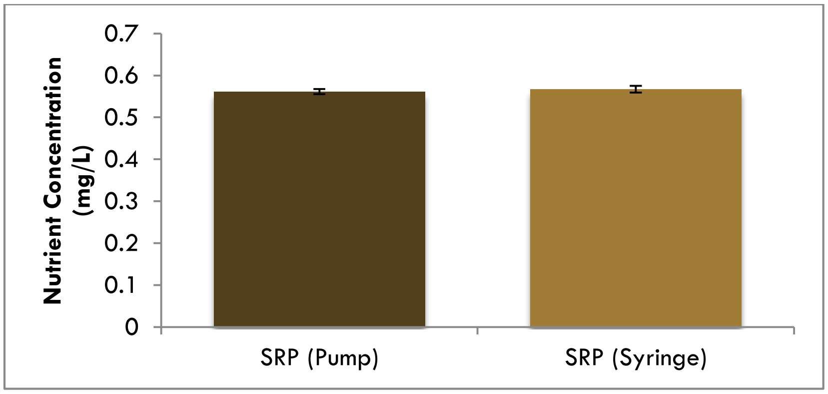


Figure 14. Comparison of soluble reactive phosphorus concentrations between samples filtered with an electric geopump and with a syringe. The error bars represent standard error between samples.

Dilution Error

Comparisons of the dilutions I made and the dilutions made in the lab showed that the lab dilutions resulted in consistently lower nutrient concentrations for every nutrient (Figures 15-19). Standard errors for the dilutions I made were also consistently higher than standard errors for the lab dilutions. This may be due to less experience with making dilutions.

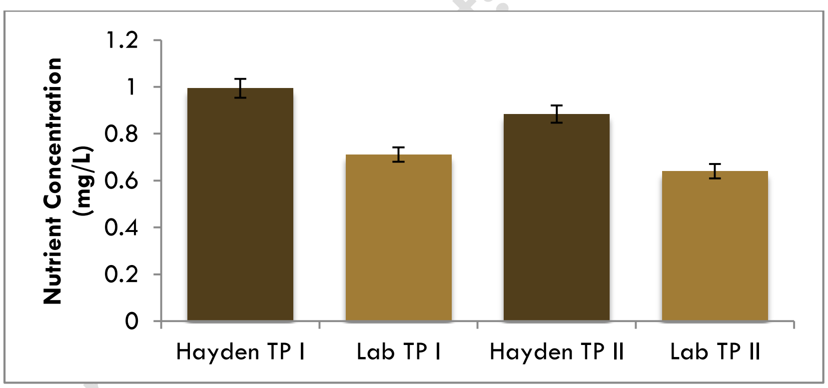


Figure 15. Comparison of total phosphorus concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

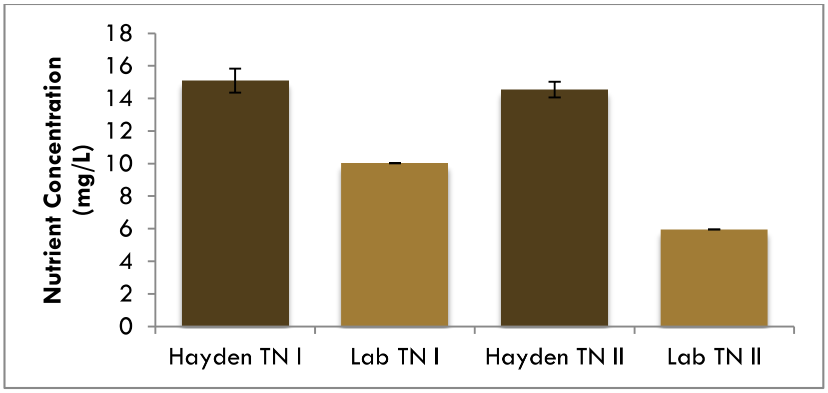


Figure 16. Comparison of total nitrogen concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

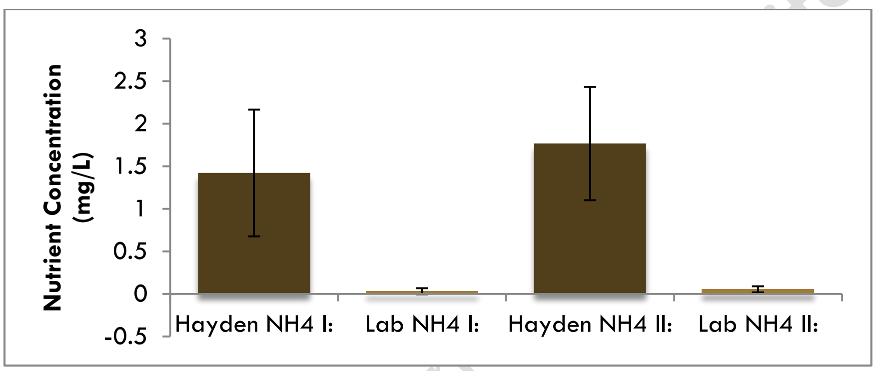


Figure 17. Comparison of ammonium concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

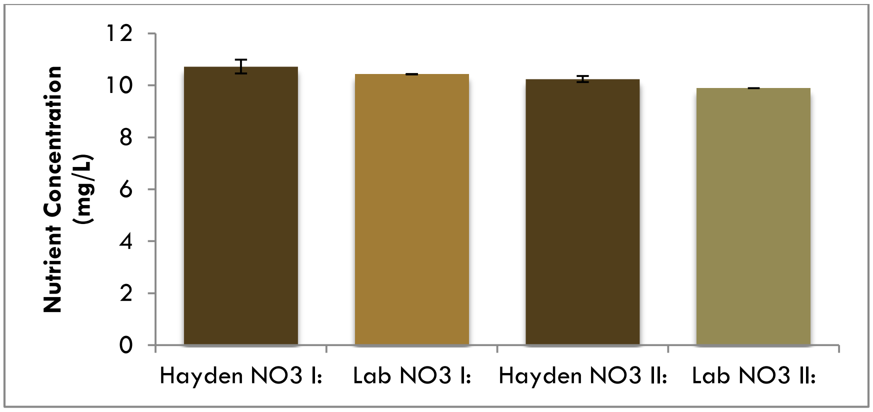


Figure 18. Comparison of nitrate concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

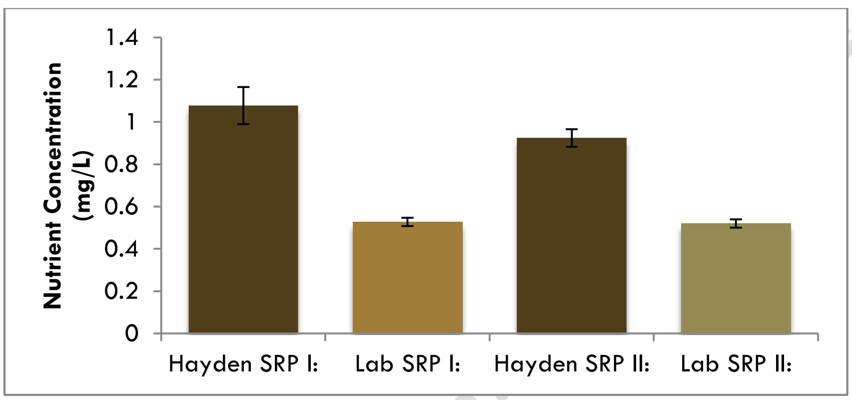


Figure 19. Comparison of soluble reactive phosphorus concentrations between samples I diluted and samples diluted in the lab. The error bars represent standard error between samples.

Analytical Uncertainty

For each nutrient, the number of fails (QA/QC data beyond 20%) were tallied for both samples spikes and duplicate samples (Tables 1 and 2). Total nitrogen and ammonium incurred the greatest number of fails, but it most cases, the percent recovery and coefficients of variations were much less than 20%.

TABLE 3. AVERAGE COEFFICIENTS OF VARIATION FOR SAMPLE DUPLICATES AND NUMBER OF FAILS ENCOUNTERED FOR EACH NUTRIENT.

	Average COV:	Number of Fails:
TN	4.2%	1
TP	4.5%	1

NH_4	-	4
NO_3	1.2%	-
SRP	6.5%	-

TABLE 4. AVERAGE PERCENT RECOVERY FOR SPIKED SAMPLES AND NUMBER OF FAILS ENCOUNTERED FOR EACH NUTRIENT.

	Average % Recovery:	Number of Fails:
TN	97.9%	3
TP	101.3%	-
NH_4	93.5%	1
NO_3	93.4%	-
SRP	98.8%	-

Storage Time

Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations ($p<0.05$) (Figures 20 and 21), however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates. The large concentration differences observed between total phosphorus concentrations of week one and the weeks following were, in large part, due to protocol failure (Figure 20). The samples analyzed on week one were locked in an autoclave overnight. The low concentrations on week one are attributed to this.

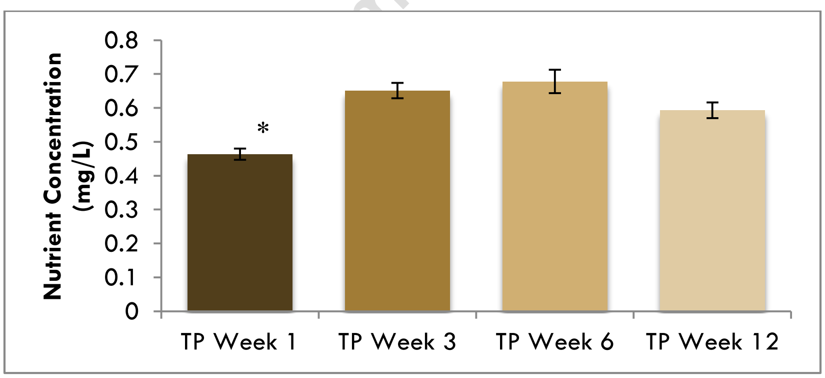


FIGURE 60. AVERAGE TOTAL PHOSPHORUS CONCENTRATIONS FOR SAMPLES ANALYZED AT ONE WEEK, THREE WEEKS, SIX WEEKS, AND TWELVE WEEKS OF FREEZING. THE ASTERISK REPRESENTS STATISTICAL SIGNIFICANCE DUE TO LAB PROTOCOL FAILURE (I.E., SAMPLES BEING LOCKED IN AN AUTOCLAVE FOR LOO LONG) AND ERROR BARS REPRESENT STANDARD ERROR BETWEEN SAMPLES.

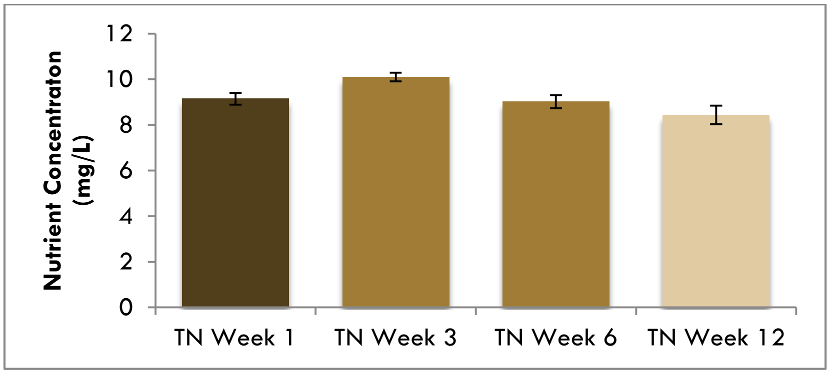


Figure 21. Average total nitrogen concentrations for samples analyzed at one week, three weeks, six weeks, and twelve weeks of freezing. The error bars represent standard error between samples.

When samples were thawed for initial analysis, put back in the freezer, and then thawed again for reanalysis, total nitrogen concentrations appeared to decrease, but total phosphorus seemed unaffected (Tables 3 and 4).

TABLE 5. TOTAL PHOSPHORUS METHOD DETECTION LIMIT AND DIFFERENCES BETWEEN SAMPLES THAWED MULTIPLE TIMES.

Total Phosphorus	mg/L
Method Detection Limit	0.43
Difference between samples opened once and opened twice	-0.007
Difference between samples opened twice and opened three times	0.03

TABLE 4. TOTAL NITROGEN METHOD DETECTION LIMIT AND DIFFERENCES BETWEEN SAMPLES THAWED MULTIPLE TIMES.

Total Nitrogen	mg/L
Method Detection Limit	0.10
Difference between samples opened once and opened twice	1.59
Difference between samples opened twice and opened three times	0.91

DISCUSSION

Nutrient mixing patterns between the four locations appeared to be different for each cross section. A general pattern was not observed; therefore mixing patterns are different for each nutrient between the four locations. According to Horowitz et al. (1990), poor selection of sampling locations within a cross-section could lead to inaccurate nutrient concentrations, due to over- or underestimation of these concentrations from variable mixing patterns. The coefficients of variation seemed to be lowest at the point source, and highest above the wastewater treatment plant, except for total nitrogen and ammonium concentrations. This could potentially lead to the hypothesis that higher concentrations confer less analytical variation; therefore concentration differences are masked by the overall high concentration.

When deciding whether to use an electric pump or a syringe when taking nutrient samples around a point source, caution should be used when analyzing for ammonium. From this data, it is not evident which filtering method is the more reliable one. When filtering with an electric pump, discretion should be used in order to assure that a tear or larger hole is not created in the filter because of too much pressure (Worsfold et al., 2005). Lambert et al. (1992) also observed the formation of filter cakes during filtration that led to changes in the effective pore size of the filter. Filters should be observed after use to determine if either of these events have occurred.

Dilutions made before analysis of nutrients were consistently lower than the dilutions I made, except for NO2+NO3-N. Variation was also greater in the dilutions I made, which is likely due to less experience. However, the differences observed between the dilutions I made and the dilutions made in the lab could be due to thawing samples multiple times before analysis. When making the dilutions, the samples were taken out of the freezer and thawed, and then placed back in the freezer. For analysis, the samples had to the thawed again. This is consistent with the data obtained from storage analysis, except total phosphorus concentrations also decreased, which was not observed in the storage analysis.

Analytical uncertainty proved to be less than uncertainty observed during sample collection and storage, except for unanticipated protocol failure. Protocol failure occurred due to samples being locked in an autoclave overnight before analysis, which caused some of the samples to completely evaporate. However, ammonium concentrations seemed to be the most variable.

Freezing samples appeared to be an adequate storage method. Samples frozen for 12 weeks showed statistically significant declines in TN and TP concentrations, however these declines were less than 9% of the initial values. This is within the range of variation seen for analytical duplicates. These results are consistent with studies done by Avanzino and Kennedy (1993), Dore et al. (1996), and Venrick and Hayward (1985). However, according to Gordolinski et al. (2001), a standard storage protocol cannot be designed due to different chemical and biological characteristics of different sample matrices. This is one reason why some studies have found freezing to be an inadequate storage technique for some nutrients. The study done by Fellman et al. (2008) is an example of a study that determined that freezing was not an adequate storage technique for total dissolved phosphorus.

According to this data, appropriate sampling methods should be used when collecting samples within a cross-section. One option is using a composite sampling technique using an automatic water sampler to get a representative sample for the whole cross section (Facchi et al., 2007; Worsford et al., 2005). Martin et al. (1992) also observed different mixing patterns due to point-source discharges, and recommended collecting grab samples at a "representative point" in a stream, if possible, or employ automatic water samplers. This may only be necessary in locations that have low nutrient concentrations, where the coefficients of variation seemed to be the greatest. Whitfield & McKinley (1981) observed that variability among field replicate samples was a great source of uncertainty in their study. This is consistent with the results of this study. Filtration methods should also be chosen appropriately when collecting samples for ammonium analysis. Finally, samples should not be thawed and frozen multiple times before analysis as this has been shown to decrease nutrient concentrations.

With these considerations in mind, the anomalies from the previous research were re-observed (Figures 22 and 23). Thawing samples multiple times could have accounted for a lower total nutrient concentration than constituent concentration if the multiple thawing only occurred in the samples analyzed for total concentrations, but not constituent concentrations. However, multiple thawing events could affect total nutrient concentrations differently than constituent concentrations, but this was not considered in this study. These anomalies could have also been due to a labeling error. The filtering technique is not a likely reason for the observed anomalies in the prior study because the anomalies are not consistent in time: samples where constituent nitrogen concentrations were higher than total nitrogen occurred early in the study, while constituent phosphorus samples were higher later in the sample collection. In addition, ammonium concentrations were low in comparison to nitrate+nitrite and total nitrogen concentrations and my research found that NH_4 could be influenced by filtration technique.

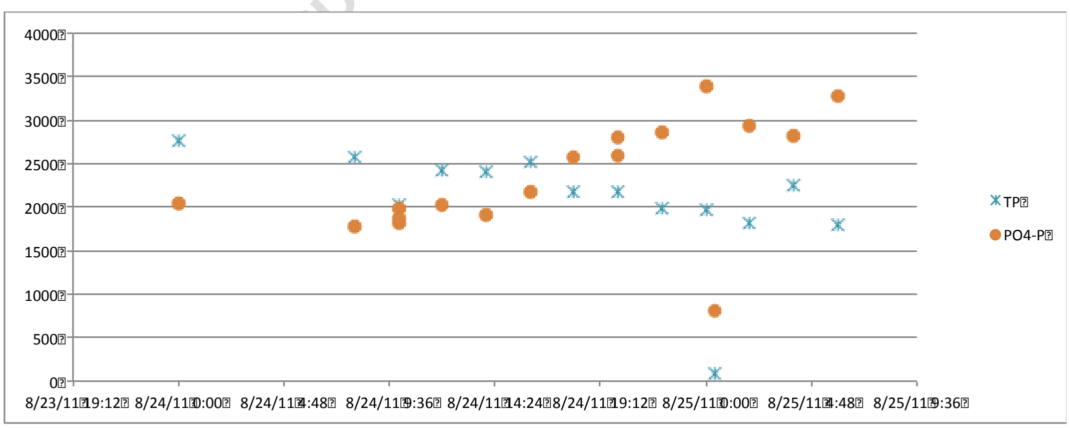


FIGURE 22. ANOMALIES OBSERVED IN PREVIOUS RESEARCH. PHOSPHATE CONCENTRATIONS WERE HIGHER THAN TOTAL PHOSPHORUS CONCENTRATIONS IN 7 OF 13 SAMPLES.

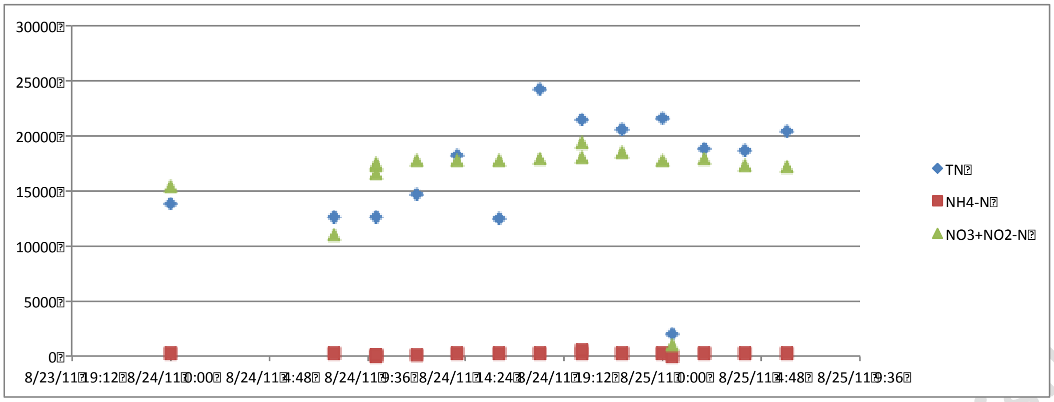


FIGURE 23. ANOMALIES OBSERVED IN PREVIOUS RESEARCH. NITRATE+NITRITE CONCENTRATIONS WERE HIGHER THAN TOTAL NITROGEN CONCENTRATIONS IN 4 OF 13 SAMPLES. HOWEVER, THE FIRST SAMPLE MAY NOT BE STATISTICALLY DIFFERENT THAN THE METHOD DETECTION LEVEL.

FUTURE RESEARCH

Thawing samples multiple times proved to decrease total nitrogen concentrations, but not total phosphorus concentrations. This research did not test whether ammonium, nitrate, or soluble reactive phosphorus concentrations also decreased as a result of multiple thaws. More research in the future is needed to test the effects of multiple thaws on these constituent nutrient concentrations. A repeat sampling event could also be performed to determine if higher nutrient concentrations do confer less analytical variation as hypothesized above.

REFERENCES

Avanzino, R. J., & Kennedy V. C. (1993). Long-term frozen storage of stream water samples for dissolved orthophosphate, nitrate Plus nitrite, and ammonia analysis. Water Resources Research, 29(10), 3357-3362.

Dore, J. E., et al. (1996) Freezing as a method of sample preservation for the analysis of dissolved inorganic nutrients in seawater. Marine Chemistry, 53, 173–185.

Facchi, A., Gandolfi C., and Whelan M. J. (2007). A Comparison of River Water Quality Sampling Methodologies Under Highly Variable Load Conditions. Chemosphere, 66(4), 746-756.

Fellman, J. B., et al. (2008). An evaluation of freezing as a preservation technique for analyzing dissolved organic C, N and P in surface water samples. Sci. Total Environ., 392(2-3), 305–312.

Horowitz, A. J., et al. (1990). Variations in suspended sediment and associated trace element concentrations in selected riverine cross sections. Environmental Science & Technology, 24(9), 1313–1320.

Jafari, A. A., & Kazemi, M. R. (2013). A parametric bootstrap approach for the equality of coefficients of variation. Comput. Stat., 28(6), 2621-2639.

Lambert, D., et al. (1992). Changes in phosphorus fractions during storage of lake water. Water Res., 26(5), 645–648.

Martin, G. R., et al. (1992). A comparison of surface-grab and cross sectionally integrated stream-water-quality sampling methods. Water Environ. Res., 64(7), 866–876.

Venrick, E. L., & Hayward T. L. (1985). Evaluation of some techniques for preserving nutrients in stored seawater samples. CalCOF Report, 26,160–168.

Whitfield, P. H., & McKinley J. W. (1981). Some factors affecting the determination of particulate carbon and nitrogen in river water. J. Am. Water Resour. As., 17(3), 381–386.

Worsfold, P., et al. (2005). Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils. Talanta, 66(2), 273–293.